

```
Ds 14 LLAALCAAGALEEKVCOGTSNRLTQLTGTFEDHFLSLQRMFNENCEVILNLEITYVQRN 73
Qy 69 ASLSFLDDIQEYGVYVLIHNOVQVPLORLRIVRGTQTLFEDNYALAVLDNGDPLNNTTP 128
Ds 74 YDLSFLKTTQEVAGYVLIHNTVERIPLENLQIRGNALYENTYALAVLSN-----124
Qy 129 VTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILKWDIFHKNQOLATLITDNR 188
Ds 125 -YGTNKTGURELPMRNLQILICAVRFSNNPLCNMETIOWRDI--QDVFLSNMSMDVQR 182
Qy 189 S-RACHPCSPMKGRGSCWSESSEDQSLTRTTCAGGCA-RCKGPLPTDCHEOCAGCTG 246
Ds 183 HLTGCPKCDPSCPNGSCWGRGECNCKLTKIICAQOCRCRCRSPSDCHNQCAAGCTG 242
Qy 247 PKHSOCLACLHNHSGICELHCPALVTYNTDTFESMNPPEGRTYTFGASCVTACPNYLST 306
Ds 243 PRESCLVCHRDEATCKDTCPLMLYNPTTYQMDVNPGEKYSFGCATCKVKCPRYVVT 302
Qy 307 DVGSCITLVCLHNLQEVTAEDGTORCEKCKPCARVCYGLGMOYIKANSKFIGITELE-PA 365
Ds 303 DHGSCURACPDYEV-EDGVSCKKCDGCRKVCNGIGIGEFK-DTISINATNIKFK 360
Qy 366 GCKKIFGSLAFSPESFDGSPASNTAPLOPEQLQVFETLEITGYLYISAWPDSLPLSVF 425
Ds 361 YCTAISGDLHLPLPAFKGDSFTRTPDLPRELEILKTVKEITGFLLIQAWPNWTDLHAF 420
Qy 426 QNLQVIRGLUNGAYSLTLOGLGISWLGSLRLSRLSGSLALIHNNTHLCFVHTVPWDL 485
Ds 421 ENLEIRGTRKQGOFSLAVVGLNITSLGSLRSLKEISDGDVITISGNRNLCYANTINWKL 480
Qy 486 RPNHQAALLHTANRDEDECVGEGLAGLCHOLCARGHCGPGTCTOCVNCQFRLGOECVEECR 545
Ds 481 FGTPNQKTKIMNRAEKCKATNHCNPLCSEGGCWGPEPTDCVSCQNVSRGREGVDKCN 540
Qy 546 VLQGLPREYVNAHCLPCHPEQOPQNGSVTCFGEADQCACAHYKDPFPCVARCPGVK 605
Ds 541 ILEGEPREFVENSECICQCHPECLPTMNITCTGRGPDNCIKAHYVDGPHCVKTCPSGM 600
Qy 606 PDLSTMPYTWKFPDEGACQPCINCTHSCVDLDDKGCAPAEQASRP-LTSIVSAVY-GILL 663
Ds 601 GENNTL-VNKFADANNVCHLCHANCTYGCAGPLAGC--QQPEGPKIPSIATGIVGGLLF 657
Qy 664 VVLGVVFGILI--QYIKANS--KFIGITEL--PLTSPGAMPNQAOMRILKTELKVK 716
Ds 658 IVVVALGIGLFMRRQLVKRTLRLQERLEVLTPSGEAPNQAHLRLKETEFKIK 717
Qy 717 VLGSAGFTGYKGIWIPGENVKIPIVAIKVLRENTSPKANKILDEAYVMAGVGPYVSR 776
Ds 718 VLGSAGFTGYKGLWIPGEKVKIPIVAIKELREATSPKANKILDEAYVMASVDNPHVCR 777
Qy 777 LLGICLTSTVQLTQMPYGCLLDHWRENRGRGLSGODLLNWCQIAKGMVSYLEDVRLVHR 836
Ds 778 LLGICLTSTVQLTQMPYGCLLDHWRENRGRGLSGODLLNWCQIAKGMVSYLEDVRLVHR 837
Qy 837 DLARNVLKSPNHVKITDFGLARLLDDETEYHADGGVKPIKMMALSESILRRRTHQSD 896
Ds 838 DLARNVLKTPQHVKITDFGLAKLGAEEKEHYHAEGGKVPKMMALSESILRIYTHQSD 897
Qy 897 VWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPCTIDVYIMVWKMWIDSE 956
Ds 898 VWSYGVTVWELMTFGSKPYDGIPIASEISSILEKGERLPQPPCTIDVYIMVWKMWIDAD 957
Qy 957 CRPRRELVSFESRNARPPQRFVWIQ-NEDLGPASPLDSTFYRSLLLEDDMGDLVDABEY 1015
Ds 958 SRPKRELILFESKWARDPQRYLVYQGDERMHLPSPTDSNFYRLMEBEDMEDVVDADBY 1017
Qy 1016 LVPOQGFCCPDPAAGAGGVHHRHSSSTRSGGDLTLGLEPSEEAEPRLAPSEGAGS 1075
Ds 1018 LIPOQGF-----NSPST-----SRTPLSLSLANS 1043
Qy 1076 DVFDGLGMGAAGKGLQSLPHTDPSPLORYSEDPVPLPSET--DGYVAPLTCSPQPEVYN 1133
Ds 1044 N-----SSTVACINRNGSCRVKEDAFQRYSDPTSLVTDNIDDTFL-----PVPEVIN 1093
```

```
Qy 1134 QPDVRRQPPSPREGPLPAARPAAGATLERAKTISPGKNVGVVQVFAFGGAVENPEYL-TPQ 1192
Ds 1094 Q-SVPRKPAQSVQNPVYHNQPLHP-----APGRDLHYQN--PHSNVSNPEYLNLTQAQ 1142
Qy 1193 CGAARQPHPPAFSPAFDNLYYWQ-----DP-----PERCAPSTFKGTPTAE 1236
Ds 1143 -----PTCLSSGFDSSALWIKQSHQMSLDNDYQODPFPKPKAKNGIFKG-PTAE 1192
Qy 1237 NPEYLGLDVP 1246
Ds 1193 NAEYLRVAPP 1202

RESULT 3
Q9EP98
ID Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mahle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schell C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mahle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275336; AAG28045.1;
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG24386.1;
DR HSSP; F11362; IFGK.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000345; CytC heme bind.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD0000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
```


SQ	SEQUENCE	1165 AA; 129614 MW; 7F7EE38D8771A74E CRC64;
Query Match	39.0%; Score 2646; DB 13; Length 1165;	
Best Local Similarity	44.7%; Pred. No. 8.7e-190;	
Matches	570; Conservative 163; Mismatches 388; Indels 154; Gaps 32;	
QY	1 MELAALCRWGLLLALLPPG-AAST-----QVCTGDMKRLRPASPETHLDMLRHLYQGCQV 55	
DB	4 LELLEL-----LLLLLSIGRCCSDPRDKVCCQTSNQMTM-----LDNHYLXMKMKYSGCV 56	
QY	56 VQGNLELTPTNASTLSFLODIQVQGVLIHNOVRQVPLQRLRIVRGTFOLFEDNYALA 115	
DB	57 VLENLEITYQENQDLSPLOSIQVGVGLIAMEVSTIPLVNLRLRQNLVYEGNFTLL 116	
QY	116 VLNDGDPPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYODTILWKDIFHK 175	
DB	117 VMSNYQK-NPSSP--DVYQVGLKQLQLSNLTEILSGGVKVSHPNLLCNVETINWMDIVDK 173	
QY	176 NNQALTLIDTNRSRACHPCSPMCKSRGWGESSEDCQSLTRTVCAGCC-ARCKGPLPTD 234	
DB	174 TSNPTWNLIPHAFERQCKCDPCGVNGSCWAPGPHQCQKFTKLLCABQCNRRCRGPID 233	
QY	235 CCEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTRGAS 294	
DB	234 CCEHCAGGCTGPRATDCLACRDNDGCTCKDTPPKIYDIYSHQVVDNPNIKYTFGA 293	
QY	295 CVTACPNYLSLTDVGSCTLYCPLHNOBVTABDGTQRCCKSPKCARVCYGLGM---QYI 350	
DB	294 CVKCEPSNVVTE-GACVRSACSAGLEV-D-ENGKRSCKPCDGVCPKCDGIGISLNTI 351	
QY	351 KANSKFIGITELEFAGACKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEEITGYL 410	
DB	352 AVNSTWIG-----SFSNCTKINGDIILNRNSFEQDPHYKIGPMDPEHLWNLTTVKEITGYL 407	
QY	411 YISAMPDLSLDVSNFQNLQVIRGRILHNGAYS-LTLQGLGSLWGLSLBELSGSLALIH 469	
DB	408 VIMWNPMTSLSVFQNLQVIRGRITTSRGSFVVQVSHLQWGLSLKEVSAGNVILK 467	
QY	470 HNTLFCVHTVPMDQLFNPHQALHTANRPEDECVGEGLAGLACHOLCARGHCWGPQTCV 529	
DB	468 NTPQLRYASTINMRLFRSEDSQSTEYDART-----ENQTCNECSEDGCGWGPPTMCV 520	
QY	530 NCQOFLRGQCEVRCVQLGLPREYNARHCLPCHPECPQONGSVTCFGEAPAOQVACAH 589	
DB	521 SCLHVRGRGCVASCNLLQGEPREAQVQDGRVCQHQCLVQDTSLTTCYGPANCSKCAH 580	
QY	590 YKDPFPCVAPCSGVKPDLSYMPYIWKFPDEEGACQPCINCTHSCVDLDDKGCPEAORAS 649	
DB	581 FQGPQCIKPRPHGLMDGDTL--IWKYADKMGQCPCHQCTQCGSGPGLSGCRGD--IVS 638	
QY	650 PLTSIVSAVVGILLVVLVGVVFGILIQYIKANSK-----FIGITEL--PLTPSGAMPNQA 702	
DB	639 HSSLAVGLVGLLITVIVALLIVLLRRRIKRRTIRRLLOEKELVEPLTPSGAPNQA 698	
QY	703 QMRILKETELRKVKVLGSGAFCTYKGIWIPDGENVKIPVAIKVLRNTPSKANKELLDE 762	
DB	699 FLRLKETEFKQDVLGSGAFCTYKGLWNPDPGENIRIPVAIKVLRNTPSKANKELLDE 758	
QY	763 AVYVAGVGSYVSRLLGICLTSVQLVTLQMPYGCLLDHRNRLGSGODLLNWCQOIA 822	
DB	759 AVYVASVDHPVCRLLGICLTSVAVQLVTLQMPYGCLLDYVRQHOERICGOWLLNWCQOIA 818	
QY	823 KGSYLEVLRVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPIKWWA 882	
DB	819 KGNVYLEERHLVHRDLAARNVLKSPNHVKITDFGLSKLLTADKEVQAHGKVPKWWA 878	
QY	883 LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 942	
DB	879 LESILQWYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLPPOPICTIE 938	
QY	943 VYMIMVKCMWIDSECRPRELSEFSRWARDPQRFVWQNEDLGPASPLDSTFYSLLE 1002	
DB	939 VTMILKCMWIDSPSRPRELSEFSQWARDPSRYLVIQG-----NLPSPSDRLFRSLLS 995	

QY	1003 DDMDGLVDABEYLVPOQGFPCPDPAQAGGMVHRRSSSTRSGGDLTLGLEPSEEA 1062	
DB	996 SDD--DVVDADEYLL-----RYKRIN-RQS----- 1018	
QY	1063 PRSLPASEGAGSDVFDGLGMGAAGKLSLTPHPPSPLOYSEDPTV-PLPSETDGYVA 1121	
DB	1019 --EPCIPNGH-----PVRENSIALRYISDPTQNALEKDLGDH-- 1054	
QY	1122 PLTCSPOPEYVNOPDVPRQP-----PSPRE-----GPLP-AARPAGATLERAKTILSPG 1168	
DB	1055 -----EYNOPGSETSSRLSDIYNPNYEDLTDGWPVSLSSQEAETNFSRPEYLNTN 1106	
QY	1169 KNGVVKDVAFAGGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYMDQDPPERCAPPST 1228	
DB	1107 QNSL---PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTGN 1141	
QY	1229 FKGTPTAENPEYLG 1243	
DB	1142 GMFLPAAENLEYLGL 1156	
RESULT 5		
Q9W6F6	PRELIMINARY; PRT; 1137 AA.	
ID	Q9W6F6	
AC	Q9W6F6;	
DT	01-NOV-1999 (TrEMBLrel. 12, Created)	
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)	
DE	Receptor tyrosine kinase (fragment).	
GN	ERBB4.	
OS	Gallus gallus (Chicken).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
OC	Gallus.	
OX	NCBI_TaxID=9031;	
RP	[1]	
RN	SEQUENCE FROM N.A.	
RC	TISSUE=HINDRAIN;	
RX	MEDLINE=99263203; PubMed=10328884;	
RA	Dixon M., Lumsden A.;	
RT	"Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in embryonic chick hindbrain."	
RL	Mol. Cell. Neurosci. 13:237-258 (1999).	
DR	EMBL; AI211963; AAD31764.1; --	
DR	HSSP; P11362; IFGK.	
DR	InterPro; IPR000494; EGFR_L_domain.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001368; TNFR_c6.	
DR	InterPro; IPR001245; Tyr_pkinase.	
DR	InterPro; IPR004019; YLP_motif.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep_L_domain; 1.	
DR	Pfam; PF02757; YLP; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; TyrKC; 1.	
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.	
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.	
DR	PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.	
KW	Kinase; Tyrosine-protein kinase.	
FT	NON_TER	
SQ	SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;	
Query Match	39.0%; Score 2643.5; DB 13; Length 1137;	
Best Local Similarity	46.2%; Pred. No. 1.3e-189;	
Matches	528; Conservative 173; Mismatches 350; Indels 91; Gaps 28;	
QY	161 LCYQDTILWKDIFHKNQLALTIDTNRSRACHPCSPMCKSGRCWGSSEDCQSLTRTV 220	


```

QY 361 ELE-FAGCKKIFGSLAFPLPSFDGDPASNTAPLOPELOQVFPETLEEITGYLYISAWPDSL 419
Db 343 NIDSFNCTKIQSLHFLVLTGILGDDFKNVPPLDAKKLEVFRTVREITDILNIQSWPEL 402
QY 420 PDLVSFQNLQVIRGRILHNGAYSITLQGLIGSWLGLRSLRGLSGLALIHNTLHLCFVHT 479
Db 403 NDLVSFSSLTIIQORSIFKRFSLMVMRIPITLSGLRSLREISDGSVYISQNAHLCHYHT 462
QY 480 VPMQDLFRNPH-QALLHTANRPEDECYGEGLACHQLCARGHCGPPTQCVCNCSQFLRGQ 538
Db 463 VNMQLFRGSRVANSLSNRPMACVADGRVCDPLCSGCGWPGDPQCLSCRNYSRHG 522
QY 539 ECVEECRVLOGLPREYNARH-CLPCHPECPQNGSVTCGPEADQCACAHYKDPFV 597
Db 523 TCVAGCHFNISGIPREFAGLNGVCVACHPECKPQTGKASCTGPGADECMACTKFRDGYCM 582
QY 598 ARCPGSGVKPDLSPYMTWKFDEGACOPCPINCTHSCVLDLDDKCPAEQASPLTSVSA 657
Db 583 SSCPAGVN-DGEKGLIFKFNREHCEPCHQNCCTQCGSGPLNDC----LEAARLTISGQ 638
QY 658 VVGILLVVLGVVF-----GILQYIKANSKFI--GITELPLTPSGAMPNOA 702
Db 639 ITGIALGVPAGLIFCLVPLFLGMLYHRGLAIRKRAMRRYLESSESFEPLGP-GEKTKV 697
QY 703 QMRILKETELRUKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 762
Db 698 HARILKPSDLRKIKLGSVGFTGSKGFWIPEGETVKIPVAIKTIQDSGSGRTFTETDH 757
QY 763 AYVMAGVGSYPVSRLLICLTSTVOLVTQMPYGCILLDHYRENRGLSGDOLLNWCVOIA 822
Db 758 LLSGSLDHYIYVRLGICPGTCLQVLTQSSHLEHROKHTSLDPORLNWCVOIA 817
QY 823 KGSYLEDVRLHRLAARNVLKSPNHVKITDPLGLARLLIDDETEYHADGKVKPIKMA 882
Db 818 KGMYLEEHRVHKLAARNILLKNDYQVQISDYGVADLLYDDKTVYSETKTPIKMA 877
QY 883 LESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPICTID 942
Db 878 LESILFRYTHQSDVMSYGVTVWEMMSFGAEPYASVQBPVPSVLEKGERLSQPAICTID 937
QY 943 VYMIMVKCMIDSECRPRELSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1002
Db 938 VYMIMVKCMIDENIRPTFKELASDFTRMARDPRLVIRMEG-----E 981
QY 1003 DDDMGDLVDAEYLVLPQGGFCPPDPAPGAGMVRHRRSSSTSSGGDLTLGLEPSEEA 1062
Db 982 DSGMGEFL-----RRCGER--GLLEADLEEDDEE- 1008
QY 1063 PRSPLAPSEGAGSDVFDGLGMC---AAKGLQSLPHDPSPLQ-----RYSEDPT 1109
Db 1009 -----GLGDRFATPSLPSPSWSTSPSQINSYVMVMTQLRYD---- 1044
QY 1110 VPLPSETDGYVAPLTCSPQ-ERYVNO-----PDVRPQPPSPREGPL--PAA 1152
Db 1045 --FAVSGQHIGYLPMSPSVDITRQLWYQSRSLSSVRLTPDRSAFRSSREAELEDGA 1102
QY 1153 RPAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPOGGAAPOPHPPPAFSPADNL 1212
Db 1103 QCAGIFRVR-----FGSERGN-----POGG----- 1122
QY 1213 YVWDQPPRGAPPTFKGTPTAENPE 1239
Db 1123 ---QQRKLSTASSPSFKTWAADDEE 1146

```

RESULT 7

```

Q9BIH9 PRELIMINARY; PRT: 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).

```

GN EGFR.

OS Anopheles gambiae (African malaria mosquito).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;

OC Anopheles.

OX NCBI_TaxID=7165;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SU4;

RA Lycett G.J.;

RT Cloning, expression and localisation of the Anopheles gambiae

RT epidermal growth factor receptor.;

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ301655; CAC35008.1; -

DR HSP; P11362; IFGK.

DR InterPro; IPR000345; CytC_heme_bind.

DR InterPro; IPR000494; EGFR_L_domain.

DR InterPro; IPR000719; Euk_pkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR002290; Ser_thr_pkinase.

DR InterPro; IPR001245; Tyr_pkinase.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF00089; pkinase; 1.

DR Pfam; PF01030; Recep_L_domain; 2.

DR PRINTS; PR00109; TYRKINASE.

DR ProDom; PD000001; Euk_pkinase; 1.

DR SMART; SM00261; FU_7.

DR SMART; SM00220; S_TKc; 1.

DR SMART; SM00219; TyrKc; 1.

DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

KW Receptor.

FT NON_TER 1 1

SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.6%; Score 2007.5; DB 5; Length 1433;

Best Local Similarity 32.3%; Pred. No. 1.2e-141;

Matches 467; Conservative 198; Mismatches 389; Indels 393; Gaps 38;

QY 26 CTCTDKMLRLPASPETHLDMLRHLYGQCQVVOGNLELTYPNALSFLQDIOBVGQYVL 85

Db 1 CIGTNGRMSVPANREYHYKRLDRYTNCTYVDGNLEITWIGNITDNLFLQHIREVTGYVL 60

QY 86 IAHQVQVPLQRLRIVRGTLQF-----EDNYALAVLDNGDPLNNTTPTVTGASPGGLREL 140

Db 61 ISYDLPQVILPRLQIIRGRTTFLKNKWEAYGLFV-----SFSHMNTL 104

QY 141 QLRSLTEILKGGVLIQRNPOLCYQDTILWKDI-FHKNNQLALTLIDTNRSPACHPCSPMC 199

Db 105 ELPALRDILGSGVGFFNNYLMCHKSNWEEILLAPOTSQYTFNFSRVCPPCHPSC 164

QY 200 KGSRCNGESSEDCQSLLTRTVACAGCA--RCKGPLPTDCCHEQCAAGCTGPKHSCLACLH 257

Db 165 EVG-OWGEAHNCQRFSLKNCSPQCSGRCFCGPKPRECCHLFCAGGCTGTQSDCLACKN 223

QY 258 FNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNYLYSTDVGSCTLVCP 317

Db 224 FYDDGVCKQECPPMQIYNPTNYFEWPNPGKYAYGATCVRKCP-EHLKKGACVRCCKP 282

QY 318 HNOEVTAEQGTQCEKSKPCARVCYGLGMOYIKANSKFGITELEFAGCKKIFGSLAF 377

Db 283 GKNPQNSE-----CVPCKGVCPTCPGEGI-----VHSDNIG-----NYKDCITIEGSL 329

QY 378 PESFDGDPASNT-----APLQPELOQVFEETGYLYISAWPDSLPLDSVFONLQ 429

Db 330 DQSGDFGQVYTNFSGPRYIKIDPDRLEVFSTVKEITGFINQAHHPNTTLYNFNLE 389

QY 430 VIRGRILHNGAY-SLTLOGLGISWLGLRSLRGLSGLALIHNTLHLCFVHTVPMDQFLRN 488

Db 390 VVGGROLKENLFAVYIVKTSLSKLSKLRVNSGIVILENSDLICFVEDIDMSKKS 449

```
Qy 489 PHQALLHTANRPEDECVEGLACHQLCARGHGWGPTQCVCNCSQFLRGQECVBEQRLVQ 548
Dy 450 SDHEVMQKRNATECHEEGECSEQCSKACWCKGPEQCLECKNVYKGLCLDSCK--- 506
Qy 549 GLPREY-VNARHCLUPCHPECPQNGSVTCFQPEADQCVACAHYKDPFCVACRP----- 601
Dy 507 SLPLYSVDSKTCGDCHQCKD-----FCYGNEDNCSCNMVXKDRGFCVACACPTTKHAM 561
Qy 602 -----SGVKPDLSPYMPWKRPD----- 618
Dy 562 NGTCINCHKTVCGRGRPRDTIAPGCCISCDKAIIGSDAKIERCLMKDESCDPGYSDYVL 621
Qy 619 -EEG----- 621
Dy 622 QEEGLQLSGKAVCRKCHPRCKTKTGYGFHEQFCQECTGYKGEQCEDECPDYANEE 681
Qy 622 --ACQCPINCT-----HSCVDL-----DD-----KGCFAEQ----- 646
Dy 682 TRICLPCHQECRGCHGLGDDHHECRNLKLFEGDPYDNATTTVCVSNCPASHPYKRFPQEA 741
Qy 647 -----RASPLTSIVSAVVGILLVVVLGWVFGILI-----QYTKA 680
Dy 742 GKIGPYCSADSMQSLRIEPTQVKIVMGVSMALILLCVFGIAFVLFSRHKNKDAVKM 801
Qy 681 NSKFIGITEL-PLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVK 739
Dy 802 TMLAGCEDSEPLRPSNVGNPLTKLRIKEAEIRRGVGLMGAFGRVFKGVWMPGESVK 861
Qy 740 IPVAIKVURENTSPANKIELDEAYVMAGVSPVSRLLGICLTSTVOLTLQMPYGLL 799
Dy 862 IPVAIKVLMEMSGSEKFELEEAYIMASVEHPHLLKLLAVCMTSQMMLLITQLMPLGCLL 921
Qy 800 DHVRENRLGSQDLNMCQIAKMSYLEDVRLVHRDLAARNVLRKSPNKHVKTDFGLA 859
Dy 922 DYVANNKDKIGSKALLNWSQTQARGMAYLEERLVHRDLAARNVLVQTPSCVKITVGLA 981
Qy 860 RLLDIDETEHADGKGVPIKMALESILRRRFTHOSDVSVMYGVTVWELMTGAKPYDGIP 919
Dy 982 KLLDFDSDEYRAAGGNPKIKWALECIRHRVFTSKSDVWAFGITIWELLTYGARPYENV 1041
Qy 920 AREIPDLLEKGERLPQPPICITIDVYIMVWKMMIDSECRPFRELVEFSFMRADPQRFV 979
Dy 1042 AKDVPELIEIGHKLPQPDICSLDVYICLLSCWLDADARPTFKQLAETFAEKARDPGRYL 1101
Qy 980 VIONEDLGPAFLDSTFYRSLLEDDDDMGDLV----- 1010
Dy 1102 MI-----PGDKFMRLPSTYNQDEKDLIRTPAVMAAAAAAAGASNDVPSTIA 1152
Qy 1011 DAEYLVPOQGFPCDPAPAGGVMVHRRHSSTRSGGDLTLGLEPSEEAAPS----- 1065
Dy 1153 ETDEYLOPKTRPSIMLPQPSA-----VEPS-DEMPKSLRYCK 1188
Qy 1066 -PLAP---SEGAGSDVFDGDLGMAAGKQLSLPTHDPSPLOQYSEDPTVPLPSETDGYVA 1121
Dy 1189 DPLAPDDETDGKGVEV-----GVGIR-----LNPLPDEDDYLM 1222
Qy 1122 PLTCSPOEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAPGG 1181
Dy 1223 P-TCOSQ-----NQS-----TPG-----YMDLIGVPA 1243
Qy 1182 AVENPEYL-----TPQGAAPQPPHPPAFSPAFDNLVYWDQDPPERGAPSTFKGT 1232
Dy 1244 SVDNPEYLMGSTOAIAGLAQSGMG--PHTPP-----PNTNPMNGM 1280
Qy 1233 PTAENPE 1239
Dy 1281 PTHQHSQ 1287
RESULT 8
Q9UK79 ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
```

```
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; PEC1BE347E2D030C CRC64;

Query Match 27.6%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 3.4e-132;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYOGCVVQGNL 60
Dy 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYOGCVVQGNL 60
Qy 61 ELYLPTNASLFLQDIQEVQGYVLIAHNOVRQVPLQRLIRVIRGTQQLFEDNYALAVLDNG 120
Dy 61 ELYLPTNASLFLQDIQEVQGYVLIAHNOVRQVPLQRLIRVIRGTQQLFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVGTGASPGGURELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNOLA 180
Dy 121 DPLNNTPTVGTGASPGGURELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTLIDTVNRACHPCSPCKGSCWGSSEDCQSLTRTVCAAGGCARCKGPLTDCCHQOC 240
Dy 181 LTLIDTVNRACHPCSPCKGSCWGSSEDCQSLTRTVCAAGGCARCKGPLTDCCHQOC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTTFESMPNPEGRYTFGASCVTACP 300
Dy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGL 345
Dy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVTHSL 345

RESULT 9
Q8R2X1 ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
```



```
Db 141 PEETATPKTGP--DHCCKAHPIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHP 197
Qy 629 NCHTSCVDLDDKCCPAEQASPLTSIVSAWV-GILLVVVLGVVFGILI--QYI---KANS 682
Db 198 NTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVGGLIGLYLRRRHIVRKRTL 254
Qy 603 KFIGITEL--PLTPSGAMPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKI 740
Db 255 RLLORELVEPLTPSGEAPNOAHLRLKETEFKKVKVLGSGAFGVYKGLMIPGEKVKI 314
Qy 741 PVAIKVLENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVQLVTQMLMPYGCLLD 800
Db 315 PVAIKELREATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQMLPYGCLLD 374
Qy 801 HVRENKRGSLGDLNWCWQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLAR 860
Db 375 YIREHKDNGSYLNCWCQIAKGNVLEERLVRDLAARNVLKTPQHVKITDFGLAK 434
Qy 861 LLDIDETEHADGGKVPKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPA 920
Db 435 LLGADEKEYHARGGKVPKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPA 494
Qy 921 RIIPDLLEKGERLPOPPICITDVMYMWKMWIDSECRPRFRELVSFSEMRARDPQFVV 980
Db 495 SEISSVLEKGERLPOPPICITDVMYMWKMWIDADSRPKFRELIAEFSKWARDPPRYLV 554
Qy 981 IQ-NEDLGPASPLDTFVRSLLDDMDGLVDAAEYLVPOQGFPCPDAPAGAGGVVHRRH 1039
Db 555 ICDERMHLPSPTDSKFRYTLMEEDMEDIVDADYLVPHQGF-----598
Qy 1040 RSSSTRSGGDLTLGLEPSEEAAPRSL-----APSEGAGSDVFDGDLGMAAKQLQSIP 1094
Db 599 NSPST-----SRTPLLSLSATSNSATNCID-----RNGQGH 632
Qy 1095 THDPSPLQRYSDPTVPLPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPA 1152
Db 633 VREDSFVQRYSSDPTGNFLEESIDGFL-----PAPEYVNO--LMPKKPS-----675
Qy 1153 RPAGATLERAKTLPQKNGVVKDF-----AFGGAIVENPEYL 1189
Db 676 -----TAMVQNIYNNISLTAISKLPMSDRYQNSHSTAVDNPEYL 715

RESULT 12
Q86714
ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhorak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
```

1- SIMILARITY: BELONGS TO THE NUCLEAR (BY SIMILARITY).
 ENBL; X52209; CAA36459.1; HSP; P10828; 2NLI.
 InterPro; IPR000719; Euk_pkinase.
 InterPro; IPR000536; Hormone_rec_lig.
 InterPro; IPR001723; Steroid_receptor.
 InterPro; IPR001245; Tyr_pkinase.
 Pfam; PF00104; hormone_rec; 1.
 Pfam; PF00069; pkinase; 1.
 PRINTS; PR00105; zf-C4; 1.
 PRINTS; PR00398; STRDHOMONER.
 ProDom; PD00047; STROIDFINGER.
 ProDom; PD000001; Euk_pkinase; 1.
 SMART; SM000035; Znf_C4steroid; 1.
 SMART; SM00430; HOLI; 1.
 SMART; SM00219; Tyrc; 1.
 PROSITE; PS00039; Znf_C4; 1.
 PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
 PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
 KW ATP-binding; DNA-binding; Nuclear_Tyr; 1.
 KW Transcription regulation; Transferase; Receptor;
 ZINC-finger.
 SQ SEQUENCE 962 AA; 108320 MW; 3CSAED791E4E95CE CRC64;

Query Match
 Best Local Similarity 23.2%; Score 1570.5; DB 15; Length 962;
 Matches 346; Conservative 75; Mismatches 145; Indels 125; Gaps 20;

541 VEERVLQGLPRE-YVNAH-HCLP-----CHPEQ 568
 354 IEKQSYLLAEHYINRYKHNIHFWKLLMKVADLRMIGAYHASRFLHMKVEPTELS 413
 569 PONGSVTCFGEADOCVACAHYKDPFCVACPCGVKPDLSYMPIWKFPDEGACQCPPI 628
 414 PQE-----VGP--DHCCKCAHFDGPHCVKACPAVLGENDTL-VKADANAVCOLCHP 465
 629 NCHTSCVDLDDKCPAEORASPLTSIVSAVV-GILLVVLGVVFGILI--OVI---KANS 682
 466 NCTRGCKGPGLEGCP---NGSKTSPSIAAGVGGILCLVVGGLGILYLRHHVVKRTLR 522
 683 KFIGITEL--PLTPSGAMPNOAOMRILKETELRKVKVLSGAGFTVYKGIWIPDGENVKI 740
 523 RLQERELVEPLTPSGEAPNOAHLRLKETEFKVKVLSGAGFTVYKGIWIPDGENVKI 740
 741 PVAIKVLRNTSPKANKELDEAYVWAGVGSPPYVRLGICLTSTVQLVLTOLMPYGCCLD 800
 583 PVAIKELREATSPKANKELDEAYVWAGVGSPPYVRLGICLTSTVQLVLTOLMPYGCCLD 800
 801 HVRENRLGSDLLNMCQIAKMSYLEDVRLVHRDLAARNVVKSPNHVKTDFGLAR 860
 643 YIREHKDNIGSYLLNMCVQIAKMSYLEDVRLVHRDLAARNVVKSPNHVKTDFGLAR 860
 861 LLDIDETEHADGKVPKMALESILRRRFTHOSDYSYGVTVWELMTFGAKPYGIPA 920
 703 QLCADKEVHAEGKVPKMALESILRRRFTHOSDYSYGVTVWELMTFGAKPYGIPA 920
 921 REIPDLLEGERLPOPPICTIDVYMWKMWIDSECPREFELVSFSEFARMARDPQRFV 980
 763 SEISSVLEKGERLPOPPICTIDVYMWKMWIDSECPREFELVSFSEFARMARDPQRFV 980
 981 IQ-NEDLGPASPLDSTYRSLLDDMGDLVDAEYLVVQQGFCDPAPGAGGVVHRH 1039
 823 IQGDERMHLPSYDSKFTLNEEDMEDIVDAEYLVVPHQGF-----866
 1040 RSSSTRSCGGDLTLGLEPSEEEAPRSLAPSEGAGSDVFDGLGMGAAGKGLSLTHDPS 1099
 867 NSPST-----SRTPLLSSLSATN-----NSATKCIDRNGCH---898

QY 1100 PLQRYSEDPVLPSETDGYVAPLTCSPQPEYVNPQDVPQPPSPREGPLPAARPA
 DB 899 -----PVREDGFL-----PAPEVYVNO--LMPKPKSTAMVQNYQIYNYI;
 QY 1159 LERAKTLSPGKNGVGVKQVFAFGGAVENPEYL 1189
 DB 938 ISKLPMDSRYQN-----SHSTAVDNPEYL 961

RESULT 14
 Q85468
 ID Q85468 PRELIMINARY;
 AC Q85468
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-MAR-2002 (TREMBLrel. 01, Last sequence update)
 DE Avian Erythroblastosis virus (Ts34) v-erbB gene.
 OS Viruses; Retrovirus; Retroviridae; Avian type C retroviruses.
 OC Viruses; Retrovirus; Retroviridae; Avian type C retroviruses.
 RN NCBI_TaxID=11861;
 RP [1]_TaxID=11861;
 RX SEQUENCE FROM N.A.
 RA MEDLINE=88217326; PubMed=28971102;
 RT "Common site of mutation in the erbB gene of avian erythroblastosis virus mutants that are temperature sensitive for transformation.";
 RL Oncogene Res. 1:265-278(1987).
 DR EMBL; X06943; CAA30024.1;
 DR HSSP; P11362; IFGK
 DR InterPro; IPR000719; Euk_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Euk_pkinase.
 DR SMART; SM000001; Euk_pkinase; 1.
 DR PROSITE; PS00107; Tyrc; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 545 AA; 60899 MW; 140DC8CCA0F8AF4 CRC64;

Query Match
 Best Local Similarity 23.1%; Score 1562; DB 15; Length 545;
 Matches 333; Conservative 72; Mismatches 124; Indels 98; Gaps 17;

578 GPRADOCVACAHYKDPFCVACPCGVKPDLSYMPIWKFPDEGACQCPINCHSCVDL 637
 1 GP--DHCCKCAHFDGPHCVKACPAVLGENDTL-VKADANAVCOLCHPCTRGCKGP 57
 638 DDKCPAEORASPLTSIVSAVV-GILLVVLGVVFGILI--OVI---KANSKFIGITEL- 690
 58 GLEGCP---NGSKTSPSIAAGVGGILCLVVGGLGILYLRHHVVKRTLRRLQERELV 114
 691 -PLTPSGAMPNOAOMRILKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLR 749
 115 EPLTPSGEAPNOAHLRLKETEFKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLR 749
 750 NTSPPKANKELDEAYVWAGVGSPPYVRLGICLTSTVQLVLTOLMPYGCCLDHHVNRGL 174
 175 ATSPKANKELDEAYVWAGVGSPPYVRLGICLTSTVQLVLTOLMPYGCCLDHHVNRGL 174
 810 GSQDLNMCQIAKMSYLEDVRLVHRDLAARNVVKSPNHVKTDFGLARLLDIDET 869
 235 GSQDLNMCQIAKMSYLEDVRLVHRDLAARNVVKSPNHVKTDFGLARLLDIDET 869
 870 HADGKVPKMALESILRRRFTHOSDYSYGVTVWELMTFGAKPYGIPASEISSVLEK 929
 295 HABGKVPKMALESILRRRFTHOSDYSYGVTVWELMTFGAKPYGIPASEISSVLEK 929
 930 GERLPOPPICTIDVYMWKMWIDSECPREFELVSFSEFARMARDPQRFVQI-NEDLGP 988
 355 GERLPOPPICTIDVYMWKMWIDSECPREFELVSFSEFARMARDPQRFVQI-NEDLGP 988

Tue Jul 22 12:42:27 2003

[illegible][illegible][illegible]

search completed: Jun-77
Job time : 51.5261 sec8

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:41:54 ; Search time 36.6984 Seconds
(without alignments)
4527.811 Million cell updates/sec

Title: SEQ4-653-675-12

Perfect score: 6776

Sequence: 1 MELALCRWGLLLALLPPCA.....TFKGTPTAENPEVLGLDVPV 1247

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6608	97.5	1255	21	Human heregulin 2
2	6608	97.5	1255	22	Human tyrosine kin
3	6608	97.5	1255	22	HER2 transgene pla
4	6608	97.5	1255	23	Human HER2 (erbB2)
5	6602	97.4	1255	17	HER-2/neu protein.
6	6602	97.4	1255	20	Human HER-2/neu on
7	6602	97.4	1255	21	Human HER-2/neu pr
8	6602	97.4	1255	21	Amino acid sequenc
9	6602	97.4	1255	22	Human HER-2/neu pr
10	6602	97.4	1255	22	HER2/neu amino aci

11	6602	97.4	1255	23	AAE24067	Human Her-2 protei
12	6602	97.4	1255	23	AAE20479	Human Her-2/neu pr
13	6602	97.4	1255	23	AAW51143	Human Her-2/neu on
14	6602	97.4	1255	23	AAU77114	Human Her-2/neu po
15	6559	96.8	1433	14	AAK39568	Sequence of c-erbB
16	6438	95.0	1223	23	AAU98923	Human breast cance
17	6285	92.8	1200	21	AAE21208	Human HER-2/neu pr
18	5819.5	85.9	1256	21	AAE21199	Rat HER-2/neu prot
19	5819.5	85.9	1256	23	AAW51144	Mouse Her-2/neu on
20	5792.5	85.5	1256	21	AAE21206	Mouse Her-2/neu pr
21	5792.5	85.5	1256	22	AAE21206	Amino acid sequenc
22	5792.5	85.5	1256	23	AAW51151	Mouse Her-2/neu on
23	4818	71.1	919	21	AAE21203	Human HER-2/neu fu
24	4818	71.1	919	23	AAW51148	Mouse Her-2/neu ex
25	4068.5	60.0	920	23	AAW51152	Mouse Her-2/neu ex
26	4068.5	60.0	926	23	AAW51153	Human HER-2/neu on
27	3702	54.6	712	21	AAE21204	Human HER-2/neu fu
28	3702	54.6	712	23	AAW51149	Her-2/neu extracel
29	3552	52.4	782	18	AAW19764	Her2-GM-CSF immuno
30	3550	52.4	653	21	AAE21200	Extracellular HER-
31	3550	52.4	653	23	AAW51145	Human Her-2/neu on
32	3512	51.8	645	22	AAE60408	Human ErbB2 oncopr
33	3512	51.8	645	22	AAE61593	Human ErbB2 extrac
34	3447	50.9	951	21	AAE44993	DC8cFV-erbB2EC fu
35	3344	49.4	624	11	AAE08222	Extracellular port
36	3066	45.2	1210	21	AAE19259	Amino acid sequenc
37	3066	45.2	1210	21	AAE50616	Human EGF receptor
38	3066	45.2	1210	23	AAE23019	Human Her-1 protei
39	3066	45.2	1210	23	AAE50768	Human epidermal gr
40	3064	45.2	1210	22	AAE68420	Amino acid sequenc
41	3053.5	45.1	654	21	AAE21205	Rat HER-2/neu prot
42	3053.5	45.1	654	23	AAW51150	Rat Her-2/neu onco
43	3025	44.6	1210	23	ABP51768	Human epidermal gr
44	2974	43.9	583	23	AAE20483	Human protein for
45	2974	43.9	587	23	AAE20481	Human protein for

ALIGNMENTS

RESULT 1
AA92620
ID AA92620 standard; Protein; 1255 AA.

XX AA92620;

XX 10-AUG-2000 (first entry)

XX Human heregulin 2 (Her2).

DE Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Domain 1..173

FT /label= N-terminal

FT /note= "mature polypeptide"

FT 5..25

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT 59..73

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT 103..117

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT 149..163

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT 174..323

```

FT Region /label= Cysteine_rich_domain
FT 210..224
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 250..264
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 324..433
FT /label= Ligand_binding_domain
FT 325..339
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 369..383
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 465..479
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 484..623
FT /label= Cysteine_rich_domain
FT 579..593
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 624..654
FT /label= Transmembrane_domain
FT 632..652
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 653..667
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 655..1010
FT /label= Tyrosine_kinase_domain
FT 661..675
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 695..709
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 710..730
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT 1011..1235
FT /label= C-terminal_domain
FT
FT WO200020027-A2.
FT
FT 13-APR-2000.
FT
FT 05-OCT-1999; 99WO-DK00525.
FT
FT 05-OCT-1998; 98DK-0001261.
FT 20-OCT-1998; 98US-0105011.
FT
FT (MEBI-) M & E BIOTECH AS.
FT
FT Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
FT Gautam A, Birk P, Karlsson G;
FT
FT WPI; 2000-349917/30.
FT N-PSDB; AAA09455.
FT
FT Inducing immune responses to weakly immunogenic, tumor associated
FT peptide antigens for the treatment of breast and prostate cancer
FT
FT Claim 62; Page 193-198; 220pp; English.
FT
FT This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
FT Her2 can be used in the claimed method as an autovaccine to induce a CTL
FT response. Subdominant CTL epitopes, antibody binding regions and
FT cysteine residues involved in disulfide bonds are preserved in the
FT immunogenized forms. Regions suitable for the insertion of foreign T
FT helper epitopes were identified (see features table). The method

```

```

CC is used for inducing immune responses against weakly immunogenic
CC cell-associated peptide antigens (PA) such as those associated with
CC cancers (self-proteins), e.g. human prostate specific membrane antigen
CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC The method comprises effecting simultaneous presentation by antigen
CC producing cells (APCs) of the animals immune system of: (1) at least 1
CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC B-cell group derived from the cell-associated PA; and (2) at least 1
CC first T helper cell group which is foreign to the animal. Analogues of
CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC part of all known and predicted CTL and B-cell epitopes of the respective
CC PA and including at least one foreign T helper epitope are also claimed.
CC The method is used to treat prostate, prostate/breast or breast cancer
CC when the PA is human PSM, FGF8b and Her2, respectively.
XX
SQ Sequence 1255 AA;
Query Match 97.5%; Score 6608; DB 21; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1222; Conservative 8; Mismatches 17; Indels 8; Gaps 3;
Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLPASPETHDMLRHLVQCQVVOGNL 60
Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLPASPETHDMLRHLVQCQVVOGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVGYVLIAHNQVQVPLQRLRIVRGTLFEDNYALVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVGYVLIAHNQVQVPLQRLRIVRGTLFEDNYALVLDNG 120
Qy 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRACHPSPCKGSRGWGESSEDCQSLTRTVAGGCARCKGPIPTDCCHQC 240
Db 181 LTLIDTNRACHPSPCKGSRGWGESSEDCQSLTRTVAGGCARCKGPIPTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACLFNHSIGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFNHSIGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMHELRVRAVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFETLSEITGYLISAWPDSL 420
Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFETLSEITGYLISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCWGPGPTQCVCNSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCWGPGPTQCVCNSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVNNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREYVNNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCPEQRASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCPEQRASPLTSIVSAVVG 660
Qy 661 ILLVVLGVVFGILV----QYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETEL 712
Db 661 ILLVVLGVVFGILVKKRQKKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 713 RKVKVLGSGAFGTGKIWIIPDGENVKIPVAIKVLRNTSPKANKELIDENYVWAGVSP 772
Db 721 RKVKVLGSGAFGTGKIWIIPDGENVKIPVAIKVLRNTSPKANKELIDENYVWAGVSP 780

```

```
QY 773 YVSRLLGICLTSTVOLVQLMPYGCILLDHYVRENRGRGLSGDILLNWCQIAKGMHSYLEYDVR 832
DB 781 YVSRLLGICLTSTVOLVQLMPYGCILLDHYVRENRGRGLSGDILLNWCQIAKGMHSYLEYDVR 840
QY 833 LVHRDLAARNVLVKSPNHVKITDFGLARLLIDIDETEHADGGKVPKIMMALESILRRRFT 892
DB 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLIDIDETEHADGGKVPKIMMALESILRRRFT 900
QY 893 HQSDVMSYGVTVWELMTFGAKPYDGI:PARBIPOLLEKGERLPPOPICTIDVYIMVKCWM 952
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGI:PARBIPOLLEKGERLPPOPICTIDVYIMVKCWM 960
QY 953 ISECEPRRELVSERWARDPQRFVIONEDLGPASPLDSTFYRSLLEDDDMGLVDA 1012
DB 961 ISECEPRRELVSERWARDPQRFVIONEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
QY 1013 EEVLVPQOQFFCPDPAPGAGGMVHHRSSRSGGDLTLGLEPSEEEAPRPLAPSEG 1072
DB 1021 EEVLVPQOQFFCPDPAPGAGGMVHHRSSRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
QY 1073 AGSDVFDGLGMAAGLQSLPHTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1132
DB 1081 AGSDVFDGLGMAAGLQSLPHTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1133 NOPDVRPQPSREGPLPAARPAATLERAKTLSPGKNGVVDVFAFGGAVENPEYLTPO 1192
DB 1141 NOPDVRPQPSREGPLPAARPAATLERAKTLSPGKNGVVDVFAFGGAVENPEYLTPO 1200
QY 1193 GGAAPOPHPPAFSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVP 1247
DB 1201 GGAAPOPHPPAFSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVP 1255
```

RESULT 2

```
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
```

```
XX Key Location/Qualifiers
FH Region 774..782
FT /note= "Antigenic epitope"
XX
XX WO200168677-A2.
XX
XX 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US40328.
XX
XX 16-MAR-2000; 2000US-0527487.
XX
XX (GENZ ) GENZYME CORP.
XX
XX Nicolette CA;
XX
XX WPI; 2001-616284/71.
XX
XX N-PSDB; AAD19731.
XX
```

Novel synthetic therapeutic compound for inducing immune response and for use in adoptive immunotherapy, has enhanced binding to major histocompatibility molecules and enhanced immunoregulatory properties

```
XX
PS
XX
CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;
```

```
Query Match 97.5%; Score 6608; DB 22; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1222; Conservative 8; Mismatches 17; Indels 8; Gaps 3;

QY 1 MELAALCRMGLLIALLPPGAASQVCTGTDMLRLPASPEHDLMLRHLVQGCQVVOGNL 60
DB 1 MELAALCRMGLLIALLPPGAASQVCTGTDMLRLPASPEHDLMLRHLVQGCQVVOGNL 60
QY 61 ELTYLPTNASLFLQDIQEVQGVLI:IAHNOVROVPLQRLIRVGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLFLQDIQEVQGVLI:IAHNOVROVPLQRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFHKNNOLA 180
DB 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFHKNNOLA 180
QY 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCCAGCARCKGPLPTDCCHEQC 240
DB 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCCAGCARCKGPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLALHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLALHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNHQVTAEDGTORCEKSKPCARVCYGLGMYKANSKFTGIT 360
DB 301 YNYLSTDVGSCTLVCPHLNHQVTAEDGTORCEKSKPCARVCYGLGMYKANSKFTGIT 360
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOQLQVFTLEETIGYLYISAWPDSL 420
DB 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOQLQVFTLEETIGYLYISAWPDSL 420
QY 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISWGLSLRELGLSGSLAIHHNTHLCFVHTV 480
DB 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISWGLSLRELGLSGSLAIHHNTHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECEGLACHQLCARGHCWGPGPTQCVNCSQFLRQEC 540
DB 481 PWDQLFRNPHQALLHTANRPEDECEGLACHQLCARGHCWGPGPTQCVNCSQFLRQEC 540
QY 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHKDPFCVCARC 600
DB 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHKDPFCVCARC 600
QY 601 PSGVKPDLISYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTSIVSAVVG 660
DB 601 PSGVKPDLISYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILII:----QYIKANS--KFIGITEL--LPTSGAMPNQAQMRILKETEL 712
DB 661 ILLVVVLGVVFGILIIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
```

```

QY 713 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 772
DB 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780
QY 773 YVSRLLGICLTSTVOLVQLMPYGCCLLDHVRENRGRIGSQDLLNMCQIAKGMYSYLEDVR 832
DB 781 YVSRLLGICLTSTVOLVQLMPYGCCLLDHVRENRGRIGSQDLLNMCQIAKGMYSYLEDVR 840
QY 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKWMALESILRRFT 892
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKWMALESILRRFT 900
QY 893 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKWM 952
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKWM 960
QY 953 IDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVA 1012
DB 961 IDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVA 1020
QY 1013 EYLVPQGFPCDPAPGAGMWHHRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1072
DB 1021 EYLVPQGFPCDPAPGAGMWHHRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
QY 1073 AGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1132
DB 1081 AGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1133 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1192
DB 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
QY 1193 GGAAPQHPHPPAPSPADNLYWQDDPPERGAPSTFKGTPTAENPEYGLDVPV 1247
DB 1201 GGAAPQHPHPPAPSPADNLYWQDDPPERGAPSTFKGTPTAENPEYGLDVPV 1255

```

RESULT 3

```

AAB60167
ID AAB60167 standard; Protein; 1255 AA.
XX
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN W0200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
XX (GETH ) GENENTECH INC.
XX
PI Erickson S, Schwall R;
XX
DR WPI; 2001-061962/07.
DR N-PSDB; AAF24297.
XX
XX Treating tumors, particularly breast cancers, which overexpress an ErbB
XX receptor and does not respond to an anti-ErbB antibody, comprises
XX conjugating the antibody to a maytansinoid -
XX

```

PS Example 3; Fig 4; 92pp; English.

XX The present invention provides a method of treating cancer by
 CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
 CC particular, the antibody is directed against ErbB2 (also known as HER2
 CC and p185neu). The method is particularly useful in the treatment of
 CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
 CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

XX Sequence 1255 AA;

```

Query Match          97.5%; Score 6608; DB 22; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1222; Conservative 8; Mismatches 17; Indels 8; Gaps 3;

QY 1 MELAALCRWGLLALLPFGAASTOVCTGTDMKRLPASPETHLMLRHLHYGCVVQGNL 60
DB 1 MELAALCRWGLLALLPFGAASTOVCTGTDMKRLPASPETHLMLRHLHYGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVGYVLIHAHQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIQEVGYVLIHAHQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGTASPGGURELQRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNQOLA 180
DB 121 DPLNNTTPTVGTASPGGURELQRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNQOLA 180
QY 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLTRTVCAAGCARCKPLPTDCCHQC 240
DB 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLTRTVCAAGCARCKPLPTDCCHQC 240
QY 241 AAGCTGPKHSDCLACLFNHSIGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFNHSIGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
DB 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
QY 361 ELEFAGCKITFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLSEITGILYISAWPDSL 420
DB 361 IQEFAGCKITFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLSEITGILYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLOGIGISWGLRSRLRELGLALIHNNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSILTLOGIGISWGLRSRLRELGLALIHNNTHLCFVHTV 480
QY 481 PWDLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGPTOCVNCSCQFLRGQEC 540
DB 481 PWDLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGPTOCVNCSCQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVAVRC 600
DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVAVRC 600
QY 601 PSGVKPDLSTYMPIWKPPDEEGACQPCINCHTSCVDLDDKGCPEAQASPLTSTVSAVVG 660
DB 601 PSGVKPDLSTYMPIWKPPDEEGACQPCINCHTSCVDLDDKGCPEAQASPLTSTVSAVVG 660
QY 661 ILLVVLGVVFGILIIKRRQKIRKYTNRRLLQETELVEPLTPSGAMPNQAMRILKETEL 712
DB 661 ILLVVLGVVFGILIIKRRQKIRKYTNRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
QY 713 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 772
DB 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780
QY 773 YVSRLLGICLTSTVOLVQLMPYGCCLLDHVRENRGRIGSQDLLNMCQIAKGMYSYLEDVR 832
DB 781 YVSRLLGICLTSTVOLVQLMPYGCCLLDHVRENRGRIGSQDLLNMCQIAKGMYSYLEDVR 840
QY 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKWMALESILRRFT 892

```

Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLIDIDETEHADGGKVPKIKMALESILRRFT 900
 QY 893 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 952
 Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
 QY 953 IDSECPRELYSEFSRMARDPQRFVVIQNERDLGPASPLDSTFYRSLLEDGMDLVDA 1012
 Db 961 IDSECPRELYSEFSRMARDPQRFVVIQNERDLGPASPLDSTFYRSLLEDGMDLVDA 1020
 QY 1013 EYLVPQOQFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1072
 Db 1021 EYLVPQOQFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 QY 1073 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1132
 Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1133 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGGAVENPEYLTPO 1192
 Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGGAVENPEYLTPO 1200
 QY 1193 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGPTTAENPEYLGLDVVPV 1247
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGPTTAENPEYLGLDVVPV 1255

RESULT 4

AAU74545

ID AAU74545 standard; Protein; 1255 AA.

XX AAU74545;

AC AAU74545;

XX AAU74545;

DT 23-APR-2002 (first entry)

XX 23-APR-2002 (first entry)

DE Human HER2 (ErbB2) polypeptide.

XX Human HER2 (ErbB2) polypeptide.

KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;

KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;

KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;

KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;

KW glial disorder; astrocytal disorder; hypothalamic disorder;

KW glandular disorder; macrophagal disorder; epithelial disorder;

KW stromal disorder; blastocoealic disorder; inflammatory disorder;

KW angiogenic disorder; immunological disorder.

KW Homo sapiens.

OS Homo sapiens.

XX Homo sapiens.

XX US2002001587-A1.

PN US2002001587-A1.

XX 03-JAN-2002.

PD 03-JAN-2002.

XX 16-MAR-2001; 2001US-0811123.

PF 16-MAR-2001; 2001US-0811123.

XX 16-MAR-2000; 2000US-189844P.

PR 16-MAR-2000; 2000US-189844P.

XX 05-OCT-2000; 2000US-238327P.

PR 05-OCT-2000; 2000US-238327P.

XX (ERIC/) ERICKSON S.

PA (SCHW/) SCHWALL R.

XX (SLIW/) SLIWKOWSKI M.

XX (SLIW/) SLIWKOWSKI M.

XX Erickson S, Schwall R, Sliwkowski M;

PI Erickson S, Schwall R, Sliwkowski M;

XX WPI; 2002-163686/21.

DR N-PSDB; ABK14058.

XX WPI; 2002-163686/21.

XX N-PSDB; ABK14058.

XX Treating tumour characterised by overexpression of epidermal growth

PT factor receptor, ErbB or cancer in mammal, comprises administering

PT anti-ErbB antibody-maytansinoid conjugate to the mammal

XX Example 3; Fig 7; 93pp; English.

PS Example 3; Fig 7; 93pp; English.

XX The invention relates to treating a tumour in a mammal, where the tumour

XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.
 XX

SQ Sequence 1255 AA;

Query Match 97.5%; Score 6608; DB 23; Length 1255;

Best Local Similarity 97.4%; Pred. No. 0;

Matches 1222; Conservative 8; Mismatches 17; Indels 8; Gaps 3;

QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPEHDLMLRHLVGGCVQVQNL 60

Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPEHDLMLRHLVGGCVQVQNL 60

QY 61 ELTYLPTNASLSFLQDIQEVQVYLAHNOVQVPLQRLRIVRGTLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIQEVQVYLAHNOVQVPLQRLRIVRGTLFEDNYALAVLDNG 120

QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRLNPOLCYQDTILWKDIFHKNOLA 180

Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRLNPOLCYQDTILWKDIFHKNOLA 180

QY 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGGCARCKGLPTDCCHQC 240

Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGGCARCKGLPTDCCHQC 240

QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMYIKANSKFIT 360

Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMYIKANSKFIT 360

QY 361 ELEFAGCKIFGSLAFELPESFDGDPASNTAPLOEQVFEETLEETIGYLIYISAWPSLP 420

Db 361 ELEFAGCKIFGSLAFELPESFDGDPASNTAPLOEQVFEETLEETIGYLIYISAWPSLP 420

QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSLRELGLSLALIHNTLHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSLRELGLSLALIHNTLHLCFVHTV 480

QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLRQEC 540

Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLRQEC 540

QY 541 VEECRVLQGLPREYVNAHCLPCHPECQONGSVTCFGEADOCVCAHYKDPFCVARC 600

Db 541 VEECRVLQGLPREYVNAHCLPCHPECQONGSVTCFGEADOCVCAHYKDPFCVARC 600

QY 601 PSGVKPDLSTYMPIWPKFPDEEGACQPCINCTHSCVDLDDKGCPEORASPLTSIVSAVVG 660

Db 601 PSGVKPDLSTYMPIWPKFPDEEGACQPCINCTHSCVDLDDKGCPEORASPLTSIVSAVVG 660

QY 661 ILLVVVLGVVFGILI----QYIKANS--KFIGITEL--PLTSGAMPNQAOMRIKTEL 712

Db 661 ILLVVVLGVVFGILI----QYIKANS--KFIGITEL--PLTSGAMPNQAOMRIKTEL 712

QY 713 RKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYVMAGVGS 772

Db 713 RKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYVMAGVGS 772

QY 721 RKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYVMAGVGS 780

Db 721 RKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYVMAGVGS 780

QY 773 YVSRLLGICLTSTVQLVTQLMPYGLLDHVRNRRGRGLSQDILLNCMQIAKMSYLEDRV 832

Db 773 YVSRLLGICLTSTVQLVTQLMPYGLLDHVRNRRGRGLSQDILLNCMQIAKMSYLEDRV 832

```
Db 781 YVSRLLGICLTSTVQLVTQMPYGCLLDHHVRENRLGSGDQLLWNCQIAKGMSTYLEDVR 840
Qy LVHRDLARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVPKWKMALESILRRFT 892
Db 841 LVHRDLARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVPKWKMALESILRRFT 900
Qy HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIMVVKWM 952
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIMVVKWM 960
Qy IDSECRPRFRELVSFESEMRARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDGDLVDA 1012
Db 961 IDSECRPRFRELVSFESEMRARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDGDLVDA 1020
Qy 1013 EYLVPQGFPCPPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1072
Db 1021 EYLVPQGFPCPPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1073 AGSDVFDGDLGMAAGKLSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1132
Db 1081 AGSDVFDGDLGMAAGKLSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1133 NOPDVRQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFATGGAVENPEYLTPO 1192
Db 1141 NOPDVRQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFATGGAVENPEYLTPO 1200
Qy 1193 GGAAPQPHPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVPV 1247
Db 1201 GGAAPQPHPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 5
AAW01111
ID AAW01111 standard; Protein; 1255 AA.
XX
AC AAW01111;
XX
DT 01-JAN-1997 (first entry)
XX
DE HER-2/neu protein.
XX
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 676..1255
FT /label= Intracellular domain
FT /note= "claimed domain, useful for immunisation"
XX
PN W09630514-A1.
XX
PD 03-OCT-1996.
XX
PF 28-MAR-1996; 96WO-US01689.
XX
PR 31-MAR-1995; 95US-0414417.
XX
PA (UNIW ) UNIV WASHINGTON.
XX
PI Cheever MA, Disis ML;
XX
DR WPI; 1996-455361/45.
DR N-PSDB; AAT40739.
XX
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
PS Claim 2; Page 56-61; 71pp; English.
```

```
XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 97.4%; Score 6602; DB 17; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;
Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLKRLPASPTHLMRLHYGCQVVGQNL 60
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLKRLPASPTHLMRLHYGCQVVGQNL 60
Qy 61 ELTYLPTNASISFLQDIOEVQGYVLI AHNOVRQVLPORLRIVRGTQLPEDNYALVLDNG 120
Db 61 ELTYLPTNASISFLQDIOEVQGYVLI AHNOVRQVLPORLRIVRGTQLPEDNYALVLDNG 120
Qy 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCAAGGCARCKGPIPTDCCHQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCAAGGCARCKGPIPTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGYRTTGASCVTACP 300
Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGYRTTGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREAVRTSAN 360
Qy 361 ELEFAGCKKIFGSLAFIPESPDGPASNTAPLQPEQLQVFTLBEITGYLYISAWPDSLP 420
Db 361 IQEFAGCKKIFGSLAFIPESPDGPASNTAPLQPEQLQVFTLBEITGYLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYS LTLQGLISWLGLRSILRELSGLALHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYS LTLQGLISWLGLRSILRELSGLALHNNTHLCFVHTV 480
Qy 481 PWDQLFRNPHOALLHTANRPEDECVBGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHOALLHTANRPEDECVBGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
Qy 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCRAEORASPLTISVAVVG 660
Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCRAEORASPLTISVAVVG 660
Qy 661 ILLVVVLGVVFGILI-----QYIKANS--KFTIGITEL--PLTPSGAMPNQAMRILKETEL 712
Db 661 ILLVVVLGVVFGILI KRRQKKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 713 RKVKVLGSGAFGTGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVWAGVSP 772
Db 721 RKVKVLGSGAFGTGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVWAGVSP 780
Qy 773 YVSRLLGICLTSTVQLVTQMPYGCLLDHHVRENRLGSGDQLLWNCQIAKGMSTYLEDVR 832
Db 781 YVSRLLGICLTSTVQLVTQMPYGCLLDHHVRENRLGSGDQLLWNCQIAKGMSTYLEDVR 840
```


QY 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 892
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
 QY 893 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKWM 952
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKWM 960
 QY 953 IDSECRPRFRELVSERMRADPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLVDA 1012
 DB 961 IDSECRPRFRELVSERMRADPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLVDA 1020
 QY 1013 EYLVPQGGFCFDPAPACAGMWHRRSSSTRSGGDLTIGLEPSEEEAPRSLAPSEG 1072
 DB 1021 EYLVPQGGFCFDPAPACAGMWHRRSSSTRSGGDLTIGLEPSEEEAPRSLAPSEG 1080
 QY 1073 AGSDVFDGDLGMAAGLQSLTHDPSPQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1132
 DB 1081 AGSDVFDGDLGMAAGLQSLTHDPSPQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1133 NQDVRPQPPSPREGPLPAARPAAGATLAKTLSPGKNGVVKOVFAFGGAVENPEYLTPO 1192
 DB 1141 NQDVRPQPPSPREGPLPAARPAAGATLAKTLSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
 QY 1193 GGAAPQHPAPAFSPAFDNLYYWDODPPERCAPPSTFKGTPTAENPEYLGLDVVP 1247
 DB 1201 GGAAPQHPAPAFSPAFDNLYYWDODPPERCAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 6

AAW92406

ID AAW92406 standard; Protein; 1255 AA.

XX AAW92406;

XX 21-APR-1999 (first entry)

XX Human HER-2/neu oncogene protein.

DE HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;

XX malignancy; treatment; tumour.

XX Homo sapiens.

XX Key Location/Qualifiers

FH Region 676..1255

FT /note= "region which elicits immune response"

XX US5869445-A.

XX 09-FEB-1999.

XX 01-APR-1996; 96US-0625101.

XX 01-APR-1996; 96US-0625101.

XX 17-MAR-1993; 93US-0033644.

XX 12-AUG-1993; 93US-0106112.

XX 31-MAR-1995; 95US-041417.

XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

XX WPI; 1999-152835/13.

XX N-PSDB; AAX01912.

XX Use of HER-2/neu polypeptides - for eliciting an immune response to

XX an HER-2/neu associated malignancy, particularly for treating or

XX preventing tumours

XX Claim 3; Column 31-38; 26pp; English.

XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or reoccurrence.

SQ Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 20; Length 1255;
 Best Local Similarity 97.2%; Pred. No. 0;
 Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;
 QY 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLPASPETHLDMLRHLGYQCQVVGNL 60
 DB 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLPASPETHLDMLRHLGYQCQVVGNL 60
 QY 61 ELTYLPTNASLSFLQDIQEVGYVLI AHNQVRQVPLQRLRIVRGTLQDPEDNYALVLDNG 120
 DB 61 ELTYLPTNASLSFLQDIQEVGYVLI AHNQVRQVPLQRLRIVRGTLQDPEDNYALVLDNG 120
 QY 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNOLA 180
 DB 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNOLA 180
 QY 181 LTLIDITNRSRACHPCSPMKGSRGWCSESSDCQSLTRTVACAGCARCKGPLTDCCHQC 240
 DB 181 LTLIDITNRSRACHPCSPMKGSRGWCSESSDCQSLTRTVACAGCARCKGPLTDCCHQC 240
 QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
 QY 301 YNYLSTDVSGCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
 DB 301 YNYLSTDVSGCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLERAVTSAN 360
 QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFPETLEETGLYLSIAPWDSLP 420
 DB 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFPETLEETGLYLSIAPWDSLP 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLSLALIHNNTHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLSLALIHNNTHLCFVHTV 480
 QY 481 PWDQFRNPHQALLHTANRPEDECVGEGACHQICARGHCWGPGPTQCVNCSQFLRGQEC 540
 DB 481 PWDQFRNPHQALLHTANRPEDECVGEGACHQICARGHCWGPGPTQCVNCSQFLRGQEC 540
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
 QY 601 PSGVKPDLSPYMPKPFDEEGACQPCPINCTHSCVDLDDKGCFAEORASPLTSIVSAVVG 660
 DB 601 PSGVKPDLSPYMPKPFDEEGACQPCPINCTHSCVDLDDKGCFAEORASPLTSIVSAVVG 660
 QY 661 ILLVVLGVVFGILI----QYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETEL 712
 DB 661 ILLVVLGVVFGILI KRRQOKIRKYMRLQLQETELVEPLTPSGAMPNQAMRILKETEL 720
 QY 713 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEITLDENYVAGVSP 772
 DB 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEITLDENYVAGVSP 780
 QY 773 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLGSDLLNWCQIAKGSYLEDVR 832
 DB 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLGSDLLNWCQIAKGSYLEDVR 840
 QY 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 892
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900

Qy 893 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWKCM 952
Db |||||||
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWKCM 960
Db |||||||
Qy 953 IDSECRPRFRELSEFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDDDMDGLVDA 1012
Db |||||||
Qy 961 IDSECRPRFRELSEFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDDDMDGLVDA 1020
Db |||||||
Qy 1013 EYLVPQGFCCPDPAFAGAGMWHRRHRSSTRSGGDLTLGLSPSEEEARSPPLAPSEG 1072
Db |||||||
Qy 1021 EYLVPQGFCCPDPAFAGAGMWHRRHRSSTRSGGDLTLGLSPSEEEARSPPLAPSEG 1080
Db |||||||
Qy 1073 AGSDVFDGLGMAAGKGLQSLTHDPSPLQRYSEDPTVPLPSETDGVYVAPLTCSPQPEYV 1132
Db |||||||
Qy 1081 AGSDVFDGLGMAAGKGLQSLTHDPSPLQRYSEDPTVPLPSETDGVYVAPLTCSPQPEYV 1140
Db |||||||
Qy 1133 NQDVRFPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGAVENPEYLTTPQ 1192
Db |||||||
Qy 1141 NQDVRFPQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVVKDVFAFGAVENPEYLTTPQ 1200
Db |||||||
Qy 1193 GGAAPQHPHPPAFSPAFDNLVWDODPPERGAPSTFKGTPTAENPEYLGLDVVP 1247
Db |||||||
Qy 1201 GGAAPQHPHPPAFSPAFDNLVWDODPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
Db |||||||

RESULT 7

ID AAB21198 standard; protein; 1255 AA.
XX AAB21198;
AC AAB21198;
DT 12-JAN-2001 (first entry)
XX Human HER-2/neu protein.
DE Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.
XX Homo sapiens.
XX WO200044899-A1.
XX 03-AUG-2000.
XX 28-JAN-2000; 2000MO-US02164.
XX 29-JAN-1999; 99US-0117976.
XX (CORI-) CORIXA CORP.
XX (SMIK) SMITHKLINE BEECHAM.
XX Cheever MA, Gheysen D;
XX WPI; 2000-505976/45.
XX N-PSDB; AAA89736.
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX Claim 52; Fig 7; 128pp; English.

CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
XX Sequence 1255 AA;
SQ Query Match
Best Local Similarity 97.4%; Score 6602; DB 21; Length 1255;
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;
Qy 1 MELAALCRWGLLLALLPFGAASSTOCTGTDMLKRLPASPTHLDMLRLHYGQCQVQGNL 60
Db 1 MELAALCRWGLLLALLPFGAASSTOCTGTDMLKRLPASPTHLDMLRLHYGQCQVQGNL 60
Qy 61 ELTYLPTNASLFLQDIQEVGYVLIHNOVQVPLQRLRIVRGTLQFEDNVALAVLNG 120
Db 61 ELTYLPTNASLFLQDIQEVGYVLIHNOVQVPLQRLRIVRGTLQFEDNVALAVLNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTELKGVLIQRPOLCYQDTILWKDIFHKNQILA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTELKGVLIQRPOLCYQDTILWKDIFHKNQILA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVAGGCARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHNEQVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHNEQVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFIPESPDGPASNTAPLQPEQLQVFTLBEITCYLYISAMPDLSL 420
Db 361 IOEFAGCKKIFGSLAFIPESPDGPASNTAPLQPEQLQVFTLBEITCYLYISAMPDLSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSRLGSLALIHNTHLFCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSRLGSLALIHNTHLFCFVHTV 480
Qy 481 PWDQLFRNPHOALLHTANRPEDECVGEGACHQLCARGHGWPGTQCVNCSOFURGQBC 540
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGACHQLCARGHGWPGTQCVNCSOFURGQBC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPECVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPECVARC 600
Qy 601 PSGVKPDLSPYMPKPFDEEGACQPCPINCCTHSCVDLDDKGPAPQASPLTSIISAVVG 660
Db 601 PSGVKPDLSPYMPKPFDEEGACQPCPINCCTHSCVDLDDKGPAPQASPLTSIISAVVG 660
Qy 661 ILLVVVLGVVFGILI---QYIKANS--KFTIGITEL--PLTPSGAMPNQAOIRILKETEL 712
Db 661 ILLVVVLGVVFGILI---QYIKANS--KFTIGITEL--PLTPSGAMPNQAOIRILKETEL 712
Qy 713 RKVKVLSGAFGTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 772
Db 721 RKVKVLSGAFGTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
Qy 773 YVSRLLGLCLTSTVQLVTQMLPYGCLLDHVRENRLGSDQLLNKCMQIAKMSYLEVDV 832
Db 781 YVSRLLGLCLTSTVQLVTQMLPYGCLLDHVRENRLGSDQLLNKCMQIAKMSYLEVDV 840
Qy 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILURRFT 892
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILURRFT 900
Qy 893 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWKCM 952
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWKCM 960
Qy 953 IDSECRPRFRELSEFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDDDMDGLVDA 1012

Db 961 IDSECRPRELVSEFSRWARDPQRFVWIONEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
QY 1013 EYLVVPQGGFFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1072
Db 1021 EYLVVPQGGFFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1073 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYV 1132
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYV 1140
QY 1133 NOPDVRPQPPSPREGPLPAARPAATLRAKTLSPGKNGVVDVFAFGGAVENPEYLTQ 1192
Db 1141 NOPDVRPQPPSPREGPLPAARPAATLRAKTLSPGKNGVVDVFAFGGAVENPEYLTQ 1200
QY 1193 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1247
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 8
AAY84780
ID AAY84780 standard; Protein; 1255 AA.
XX
AC AAY84780;
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX
KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.
XX
OS Homo sapiens.
XX
PN WO200020579-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00912.
XX
PR 02-OCT-1998; 98US-0165192.
XX
(UYMC-) UNIV MCMASTER.
XX
PI Muller WJ, Siegel PM;
XX
DR WPI; 2000-303768/26.
XX
DR N-PSDB; AAA14812.
PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.
CC The present sequence represents a SPLICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
SQ Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 21; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;
QY 1 MELAAALCRWGLLLALLPPGNAASTQVCTGTDMLKRLPASPTHLDMLRHLVQGCQVOVGNL 60
Db 1 MELAAALCRWGLLLALLPPGNAASTQVCTGTDMLKRLPASPTHLDMLRHLVQGCQVOVGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVVLAHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVVLAHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFHKNNQLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFHKNNQLA 180
QY 181 LTLIDNTRSRACHPCSPMKCGSRGWESSEDCOSLTRTVCCAGGCARCKGPLPTDCCHQEC 240
Db 181 LTLIDNTRSRACHPCSPMKCGSRGWESSEDCOSLTRTVCCAGGCARCKGPLPTDCCHQEC 240
QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQCEKCKSPCARVCYGLGMYIKANSKFITGIT 360
Db 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQCEKCKSPCARVCYGLGMEHLREVRAVTSAN 360
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVPETLEEITGYIYISAWPDSLP 420
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVPETLEEITGYIYISAWPDSLP 420
QY 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISMGLRSLRELGSGLALIHNTHLFCFVHTV 480
Db 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISMGLRSLRELGSGLALIHNTHLFCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVNCSOFLRQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVNCSOFLRQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECQONGSVTCFGEADQCVACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECQONGSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 601 PSGVKPDLSYMPIWKFPEDEGACQPCINCTHSCVDLDDKGCPEAQBASPLTSIVSAVVG 660
Db 601 PSGVKPDLSYMPIWKFPEDEGACQPCINCTHSCVDLDDKGCPEAQBASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILI---QYIKANS--KFIGITEL--PLTPSGAMPNOQMRLKETEL 712
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNOQMRLKETEL 720
QY 713 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKSILDEAYVMAGVGP 772
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKSILDEAYVMAGVGP 780
QY 773 YVSRLLGICLTSTVOLVTOLMPYGCLLDHVRENRLGSGQDLLNWCQIAGKSYLEDVR 832
Db 781 YVSRLLGICLTSTVOLVTOLMPYGCLLDHVRENRLGSGQDLLNWCQIAGKSYLEDVR 840
QY 833 LVHRLDAARNLVKSPNKHVITDFGLARLLDIDETEHADGGKVPKKNWALESLRRRFT 892
Db 841 LVHRLDAARNLVKSPNKHVITDFGLARLLDIDETEHADGGKVPKKNWALESLRRRFT 900
QY 893 HQSDVMSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITIDVYIMVKCWM 952
Db 901 HQSDVMSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITIDVYIMVKCWM 960
QY 953 IDSECRPRELVSEFSRWARDPQRFVWIONEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1012
Db 961 IDSECRPRELVSEFSRWARDPQRFVWIONEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
QY 1013 EYLVVPQGGFFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1072

```
Db 1021 EEVLVPOQGFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080
QY 1073 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1132
Db 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1133 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVVDVFAFGGAVENPEYLTPO 1192
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVVDVFAFGGAVENPEYLTPO 1200
QY 1193 GGAAPQHPPPAFSPAFDNLYYDQDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1247
Db 1201 GGAAPQHPPPAFSPAFDNLYYDQDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX
AC AAB85458;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
OS Homo sapiens.
XX
PN W0200153463-A2.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001WO-US01850.
XX
PR 21-JAN-2000; 2000US-0177545.
XX
PA (CORI-) CORIXA CORP.
XX
PI Cheever MA, Hand-Zimmermann S;
XX
PI WPI; 2001-476112/51.
XX
DR N-PSDB; AAH23392.
XX
PT New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer -
XX
XX Claim 2; Page 41-46; 49pp; English.
XX
CC The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 22; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;

QY 1 MEAAACRWGLLALLPPGAASQVCTGTDKMLRLPASPETHLDMLRHLYQGCGVQGNL 60
Db 1 MEAAACRWGLLALLPPGAASQVCTGTDKMLRLPASPETHLDMLRHLYQGCGVQGNL 60
QY 61 ELTYLPTNASLFLQDIEQVGVVLIHAHQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
```

```
Db 61 ELTYLPTNASLFLQDIEQVGVVLIHAHQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNOLA 180
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNOLA 180
QY 181 LTLIDNRSRACHPCS PMCKGSCWGESSEDCOSLTRTVCCAGGCARCKGPLPTCCCHQEC 240
Db 181 LTLIDNRSRACHPCS PMCKGSCWGESSEDCOSLTRTVCCAGGCARCKGPLPTCCCHQEC 240
QY 241 AAGCTGPKGSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKGSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRRAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRRAVTSAN 360
QY 361 ELEPAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITCYLYVISAWPDSL 420
Db 361 IQEPAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITCYLYVISAWPDSL 420
QY 421 DLSVFQNLQVIRGRIHLHGAYSLTLQGLGISWLGRLSLRELGLGLALIHNTHLFCFVHTV 480
Db 421 DLSVFQNLQVIRGRIHLHGAYSLTLQGLGISWLGRLSLRELGLGLALIHNTHLFCFVHTV 480
QY 481 PWDLFRNPHOALLHTANRPEDEBCVGBGLACHQICARGHCWGPPTQCVNCSQFLRGQEC 540
Db 481 PWDLFRNPHOALLHTANRPEDEBCVGBGLACHQICARGHCWGPPTQCVNCSQFLRGQEC 540
QY 541 VEESRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Db 541 VEESRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
QY 601 PSGVKPDLSTYMPIWKFPDEEGACQPCINCTHSCVDLDDKCPAEQASPLTSTVSAVVG 660
Db 601 PSGVKPDLSTYMPIWKFPDEEGACQPCINCTHSCVDLDDKCPAEQASPLTSTVSAVVG 660
QY 661 ILLVVVLGVVFGILJ----QYIKANS--KFIGITEL--PLTPSGAMPNQAQMRILKETEL 712
Db 661 ILLVVVLGVVFGILJIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 713 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKETILDEAYVMAGVCS 772
Db 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKETILDEAYVMAGVCS 780
QY 773 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGRIGSQDLLNWCQIAKGMSTYLEDVR 832
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGRIGSQDLLNWCQIAKGMSTYLEDVR 840
QY 833 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRFT 892
Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRFT 900
QY 893 HQSDVMSYGVTVVELMTFFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVKCM 952
Db 901 HQSDVMSYGVTVVELMTFFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVKCM 960
QY 953 IDSECRPRFRELVESEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDGLVDA 1012
Db 961 IDSECRPRFRELVESEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDGLVDA 1020
QY 1013 EYLVPOQGFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1072
Db 1021 EYLVPOQGFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080
QY 1073 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1132
Db 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1133 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVVDVFAFGGAVENPEYLTPO 1192
```

Db 1141 NQPDVTRPQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVVDVAFAGGAVENPEYLTPQ 1200
QY 1193 GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPPEFTFKGTTPTAENPEYLGDLVVP 1247
Db 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPPEFTFKGTTPTAENPEYLGDLVVP 1255

RESULT 10

AAG88267

ID AAG88267 standard; Protein; 1255 AA.

XX AAG88267;

DT 11-SEP-2001 (first entry)

XX HER2/neu amino acid sequence.

XX Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
XX immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
XX tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.

XX Homo sapiens.

XX WO200141787-A1.

XX 14-JUN-2001.

XX 11-DEC-2000; 2000WO-US33591.

XX 10-DEC-1999; 99US-0458299.

XX (EPTM-) EPIMUNE INC.

XX Pikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;

XX Keogh E;

XX WPI; 2001-374995/39.

XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
XX cellular immune responses for the prevention and treatment of cancer -
XX Disclosure; Page 15; 199pp; English.

XX The present invention describes isolated prepared HER2/neu epitopes (I).
XX Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
XX culture in vitro and binds to a complex of an epitope (I), bound to a
XX human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
XX and a second epitope and the peptide is less than 50 contiguous amino
XX acids that have 100% identity with a native peptide sequence of HER2/neu;
XX (3) a vaccine composition (III) comprising (II) and a pharmaceutical
XX excipient; (4) an isolated nucleic acid encoding a peptide comprising
XX (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
XX and immunostimulant activities, and can be used in vaccines. (I), (II)
XX and (III) are useful for inducing cellular immune responses for the
XX prevention and treatment of cancer. (I) and (II) are useful for
XX monitoring or evaluating an immune response to a tumour-associated
XX antigen when incubated with a T lymphocyte sample from a patient and
XX detecting the presence of bound T lymphocyte to (I) or (II). Epitope
XX based vaccines mean that immunosuppressive epitopes that may be present
XX in whole antigens may be avoided. Selected epitopes may be combined to
XX enhance immunogenicity. The possible pathological side effects caused by
XX infectious agents or whole protein antigen is eliminated. The vaccine
XX provides the ability to direct and focus an immune response to multiple
XX selected antigens from the same pathogen. Epitope-based anti-tumour
XX vaccines provides the opportunity to combine epitopes derived from
XX multiple tumour-associated molecules addressing the problem of tumour-
XX tumour variability and reducing the likelihood of tumour escape due to
XX antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
XX the exemplification of the present invention.

XX Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 22; Length 1255;

Best Local Similarity 97.2%; Pred. No. 0;
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;

QY	1	MELAAALCRGLLALLPPGAAS	TQVCTGTDMKLR	LSPAS	PTH	LDMLRHL	YOGC	VOVGNL	60			
Db	1	MELAAALCRGLLALLPPGAAS	TQVCTGTDMKLR	LSPAS	PTH	LDMLRHL	YOGC	VOVGNL	60			
QY	61	ELTYLPTNASLSFLQDIQ	EVQGVLI	AHNOVR	QVPLQRL	RIVRGT	QFED	NYALAVLDNG	120			
Db	61	ELTYLPTNASLSFLQDIQ	EVQGVLI	AHNOVR	QVPLQRL	RIVRGT	QFED	NYALAVLDNG	120			
QY	121	DPLNNTTPTVTGASPGGL	RELQRLSL	TEILKGG	VLIO	RNPOL	CYOD	TILWKDIFHKNQOLA	180			
Db	121	DPLNNTTPTVTGASPGGL	RELQRLSL	TEILKGG	VLIO	RNPOL	CYOD	TILWKDIFHKNQOLA	180			
QY	181	LTLIDTNRSPACHPCSP	CMCKSG	RCWGES	SDCOSL	TRTV	CAGG	CARCKG	PLTDCCHQC	240		
Db	181	LTLIDTNRSPACHPCSP	CMCKSG	RCWGES	SDCOSL	TRTV	CAGG	CARCKG	PLTDCCHQC	240		
QY	241	AAGCTGPKHSDCLACL	HFHNSG	ICELH	CPAL	VTYNTD	TFS	MPNPEGR	YTFGASCVTACP	300		
Db	241	AAGCTGPKHSDCLACL	HFHNSG	ICELH	CPAL	VTYNTD	TFS	MPNPEGR	YTFGASCVTACP	300		
QY	301	YNYLSTDVGSCTLV	CPHLNQ	EVTAED	GTQRC	CKSP	CARVC	YGLG	MOYIKANSK	FGIT	360	
Db	301	YNYLSTDVGSCTLV	CPHLNQ	EVTAED	GTQRC	CKSP	CARVC	YGLG	MEHLRE	VAVTSAN	360	
QY	361	ELFEGACKI	FGSLAF	LPES	FDGDP	ASNTAP	LOPE	QLOV	FETLEE	ITGYL	ISAWP	420
Db	361	IOEFAGCKI	IFGSLAF	LPES	FDGDP	ASNTAP	LOPE	QLOV	FETLEE	ITGYL	ISAWP	420
QY	421	DLVSFQNLQVIRGR	ILHNGAYS	SLT	LOGLG	ISWGL	SLRSL	REL	SGSLAL	IHNH	THLCF	480
Db	421	DLVSFQNLQVIRGR	ILHNGAYS	SLT	LOGLG	ISWGL	SLRSL	REL	SGSLAL	IHNH	THLCF	480
QY	481	PMDQLFRNPHQALL	HTANRPE	DEC	VGEG	LACH	OLCARG	HCG	PGPTQ	CVNC	SQFLRQ	540
Db	481	PMDQLFRNPHQALL	HTANRPE	DEC	VGEG	LACH	OLCARG	HCG	PGPTQ	CVNC	SQFLRQ	540
QY	541	VVEECRVQLG	LPREYNAR	HCLPCH	PECOP	QNGSV	TCF	GP	PEADOC	VACAH	KDPF	600
Db	541	VVEECRVQLG	LPREYNAR	HCLPCH	PECOP	QNGSV	TCF	GP	PEADOC	VACAH	KDPF	600
QY	601	PSGVKPDLS	YMPIWKP	PEEG	ACQPC	PIN	CTH	SCVD	LDKGC	PAEOR	ASPLT	660
Db	601	PSGVKPDLS	YMPIWKP	PEEG	ACQPC	PIN	CTH	SCVD	LDKGC	PAEOR	ASPLT	660
QY	661	ILLVVVLGVV	VGIL	---	QYIKANS	--	XF	IGITEL	--	PL	TPSG	712
Db	661	ILLVVVLGVV	VGIL	---	QYIKANS	--	XF	IGITEL	--	PL	TPSG	712
QY	713	RKVVLGSG	ARGTVY	KGIW	IPDGEN	VKI	PVA	IKV	RENT	SP	KANK	772
Db	721	RKVVLGSG	ARGTVY	KGIW	IPDGEN	VKI	PVA	IKV	RENT	SP	KANK	772
QY	773	YYSRLIG	ICLT	STVQ	LVLT	QMP	YVCL	LDH	VRENR	GR	LSG	832
Db	781	YYSRLIG	ICLT	STVQ	LVLT	QMP	YVCL	LDH	VRENR	GR	LSG	832
QY	833	LVHRDLA	ARNV	LK	SPNH	VKIT	DF	GLAR	LLD	ID	ET	892
Db	841	LVHRDLA	ARNV	LK	SPNH	VKIT	DF	GLAR	LLD	ID	ET	892
QY	893	HQSDVMS	YGV	TW	ELMT	FGAK	PDY	GI	PARE	I	PD	952
Db	901	HQSDVMS	YGV	TW	ELMT	FGAK	PDY	GI	PARE	I	PD	952
QY	953	IDSECR	PR	PRE	LVS	EF	SR	MR	DP	QRF	W	1012
Db	961	IDSECR	PR	PRE	LVS	EF	SR	MR	DP	QRF	W	1012
QY	1013	EYLV	VP	QQ	FF	CP	D	P	A	G	G	1072
Db	1013	EYLV	VP	QQ	FF	CP	D	P	A	G	G	1072

Db 1021 EELVLPQOQFFCDDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 Qy 1073 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYV 1132
 Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYV 1140
 Qy 1133 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTQ 1192
 Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVVKDVFAFGGAVENPEYLTQ 1200
 Qy 1193 GGAAPQHPPPAPSPAFDNLVYWDQPPPERGAPPTFKGPTTAENPEYLGLDVVP 1247
 Db 1201 GGAAPQHPPPAPSPAFDNLVYWDQPPPERGAPPTFKGPTTAENPEYLGLDVVP 1255

RESULT 11

AAE24067
 ID AAE24067 standard; Protein; 1255 AA.

XX AC AAE24067;
 DT 23-SEP-2002 (first entry)
 DE Human Her-2 protein.
 XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
 KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
 KW tumour; gene therapy; phosphorothioate backbone.

XX Homo sapiens.

OS WO200222636-A1.

XX 21-MAR-2002.

XX 12-SEP-2001; 2001WO-US28572.

XX 15-SEP-2000; 2000US-0663834.

XX (ISIS-) ISIS PHARM INC.

PA Bennett CF, Cowser LM;

PI WPI; 2002-471192/50.

DR N-PSDB; AAD38904.

XX Novel antisense oligonucleotide which modulates the expression of Human
 PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
 PT inflammation or to prevent infection in humans -

XX Example 13; Page 95-107; 116pp; English.

XX The invention relates to antisense compounds targetted to a nucleic
 CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
 CC that specifically hybridises with and inhibits the expression of Her2.
 CC Antisense compounds of the invention are used for treating diseases or
 CC conditions associated with Her2 such as hyperproliferative disorders
 CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
 CC neural or cardiac cancer. They are also useful prophylactically e.g.
 CC to prevent or delay infection, inflammation and tumour formation. The
 CC invention is also used in gene therapy. The present sequence is human
 CC Her-2 protein.

XX Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 23; Length 1255;

Best Local Similarity 97.2%; Pred. No. 0;

Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPTHLDMLRHLYQGCVVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPTHLDMLRHLYQGCVVQGNL 60

Qy 61 ELTYLPTNASLFLQDIQEVQGYVLIHNOVROVPLQRLRIVRGTOLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLFLQDIQEVQGYVLIHNOVROVPLQRLRIVRGTOLFEDNYALAVLDNG 120
 Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDI FHKNNOLA 180
 Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDI FHKNNOLA 180
 Qy 181 LTLIDNRSRACHPCSPMKCKSGCWESSEDCOSLRTVTCAGGCARCKGPLPTDCCHEQC 240
 Db 181 LTLIDNRSRACHPCSPMKCKSGCWESSEDCOSLRTVTCAGGCARCKGPLPTDCCHEQC 240
 Qy 241 AAGCTGPKISDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKISDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPLNHQEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIT 360
 Db 301 YNYLSTDVGSCTLVCPLNHQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
 Qy 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLOVFETLEEITGYLYISAMPDLSL 420
 Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLOVFETLEEITGYLYISAMPDLSL 420
 Qy 421 DLSVFQNLQVIRGRILHNGAYSILTLOGLGISWLGSLRLSGSLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSILTLOGLGISWLGSLRLSGSLALIHNTLHLCFVHTV 480
 Qy 481 PWDQLFRNHQALLHTANRPEDECVGEGLAHQCHGCHGCPGPTQCVNCSQPLRQEC 540
 Db 481 PWDQLFRNHQALLHTANRPEDECVGEGLAHQCHGCHGCPGPTQCVNCSQPLRQEC 540
 Qy 541 VEESCRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHKDPFPCVARC 600
 Db 541 VEESCRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHKDPFPCVARC 600
 Qy 601 PSGVKPDLSPYMPIWKFPDEBACQPCINCTHSCVDLDDKCPAEORASPLTSISAVVG 660
 Db 601 PSGVKPDLSPYMPIWKFPDEBACQPCINCTHSCVDLDDKCPAEORASPLTSISAVVG 660
 Qy 661 ILLVVLGVVFGILI---QYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETEL 712
 Db 661 ILLVVLGVVFGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Qy 713 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 772
 Db 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
 Qy 773 YVSRLLGICLTSTVQLVTOLMPYGCLLDHVRENRLGSGDQLLNWCNMQIAKMSYLEDVR 832
 Db 781 YVSRLLGICLTSTVQLVTOLMPYGCLLDHVRENRLGSGDQLLNWCNMQIAKMSYLEDVR 840
 Qy 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRFT 892
 Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRFT 900
 Qy 893 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICITIDVYIMVVKCW 952
 Db 901 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICITIDVYIMVVKCW 960
 Qy 953 IDSECRPRFRELVSFERSWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDGLVDA 1012
 Db 961 IDSECRPRFRELVSFERSWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDGLVDA 1020
 Qy 1013 EELVFPQOQFFCDDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1072
 Db 1021 EELVFPQOQFFCDDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 Qy 1073 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYV 1132
 Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYV 1140
 Qy 1133 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTQ 1192

Db 1141 NOPDVRPQPPSPREGPLPAARAGATLERPKTSLSGKNGVVKDVFAGGAVENPEYLTTPQ 1200
QY 1193 GGAAPQHPHPPAFSPAFDNLYYWDQDPPRGAPPSTFKGTPTAENPEYILGLDVPV 1247
Db 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPRGAPPSTFKGTPTAENPEYILGLDVPV 1255

RESULT 12
ID AAE20479
XX AAE20479 standard; Protein; 1255 AA.
AC AAE20479;
DT 01-JUL-2002 (first entry)
XX Human Her-2/neu protein.
DE Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX Homo sapiens.

OS
FH Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
FT
FN WO200214503-A2.

PN 21-FEB-2002.
XX 14-AUG-2001; 2001WO-US41733.
XX 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-236428P.
PR 21-FEB-2001; 2001US-270520P.
XX (CORI-) CORIXA CORP.
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI Mcneill PD, Vedvick TS;
XX WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer
XX Disclosure; Page 114-117; 129pp; English.

XX The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.

XX Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 23; Length 1255;

Best Local Similarity 97.2%; Pred. No. 0;
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;
QY 1 MELAALCRWGLLALLPPGAASCTGCTDMKLRLPASPETHLDMRLHLHYOGCOVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASCTGCTDMKLRLPASPETHLDMRLHLHYOGCOVQGNL 60
QY 61 ELYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLIRVGTOLFDENYALAVLDNG 120
Db 61 ELYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLIRVGTOLFDENYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPCKGSRGSESSDCQSLTRTYCAGCARGCKPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPCKGSRGSESSDCQSLTRTYCAGCARGCKPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNHNVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPLNHNVTAEDGTORCEKSKPCARVCYGLGMEHLREAVTSAN 360
QY 361 ELEFAGCKKIFGSLAFLPESPDGPASNTAPLOPELOQVFEETLEETIGLYISAWPDSL 420
Db 361 IQEFAGCKKIFGSLAFLPESPDGPASNTAPLOPELOQVFEETLEETIGLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRIILHNGAYSLTLQGLGISWLGRLSRELGSGLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRIILHNGAYSLTLQGLGISWLGRLSRELGSGLALIHNNTHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNRHCLPCHPECOPOGNSVTCFGEADOCVACAHYKPPFCVARC 600
Db 541 VEECRVLQGLPREYVNRHCLPCHPECOPOGNSVTCFGEADOCVACAHYKPPFCVARC 600
QY 601 PSGVKPDLSPYMPIWKFDEEGACQPCINCTHSCVDLDDKGCAPAEORASPLTSIYSAVVG 660
Db 601 PSGVKPDLSPYMPIWKFDEEGACQPCINCTHSCVDLDDKGCAPAEORASPLTSIYSAVVG 660
QY 661 ILLVVVLGVVFGILII---QYIKANS--KFIGITEL--PLTPSGAMPNQAQMRILKETEL 712
Db 661 ILLVVVLGVVFGILIIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 713 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGS 772
Db 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGS 780
QY 773 YVSRLLGICLTSTVQLVTQLMPYGLLDHVRNRLGSDLLNWCMIKAGMSYLEDVR 832
Db 781 YVSRLLGICLTSTVQLVTQLMPYGLLDHVRNRLGSDLLNWCMIKAGMSYLEDVR 840
QY 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRT 892
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRT 900
QY 893 HQSDWVSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICTIDVYMIKVCWM 952
Db 901 HQSDWVSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICTIDVYMIKVCWM 960
QY 953 IDSECRPRFRELVSFSEFRMARDPQRFVVIQNEIDLGASPLDSTFYRSILLEDDMDGLVDA 1012
Db 961 IDSECRPRFRELVSFSEFRMARDPQRFVVIQNEIDLGASPLDSTFYRSILLEDDMDGLVDA 1020
QY 1013 EYLVPOGGFFCPDPAFGAGMVRHRSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1072

Db 1021 BEYLVPOQGFCCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEB 1080
Qy 1073 AGSDVFDGLGMAAKGLOSLTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1132
Db 1081 AGSDVFDGLGMAAKGLOSLTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1133 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLPCKNGVWVDVFAFGGAVENPEYLTPO 1192
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERPKTLPCKNGVWVDVFAFGGAVENPEYLTPO 1200
Qy 1193 GGAARQPHPPAFSPAFDNLNLYWDQPPRGAPPSTFKGTPTAENPEYLGDLVVPV 1247
Db 1201 GGAARQPHPPAFSPAFDNLNLYWDQPPRGAPPSTFKGTPTAENPEYLGDLVVPV 1255

RESULT 13
AAMS1143
ID AAMS1143 standard; Protein; 1255 AA.
AC AAMS1143;
XX
DT 17-JUN-2002 (first entry)
XX Human Her-2/neu oncogene-encoded p185 glycoprotein.
DE Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KW tyrosine kinase; receptor; c-erbB2; gene therapy.
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT Domain 1..653
FT /note= "extracellular domain"
FT Domain 676..1255
FT /note= "intracellular domain"
FT Domain 990..1255
FT /note= "phosphorylation domain"
FT
XX WO200212341-A2.
PN
XX
PD 14-FEB-2002.
XX
PF 03-AUG-2001; 2001WO-US24283.
XX
PR 03-AUG-2000; 2000US-0632507.
XX
PA (CORI-) CORIXA CORP.
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Cheever MA, Gheyssen D;
XX
DR WPI; 2002-241743/29.
DR N-PSDB; ABA92250.
XX
PT Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT or enhancing an immune response to the protein, has Her-2/neu
PT extracellular domain fused to Her-2/neu intracellular or
PT phosphorylation domain -
XX
PS Claim 68; Fig 7; 141pp; English.
XX
CC The present sequence is that of human Her-2/neu (p185 glycoprotein
CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
CC in a variety of cancers, including breast, ovarian, colon, lung and
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
CC family of receptor-like glycoproteins. It comprises an extracellular
CC domain with homology to the epidermal growth factor receptor
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC intracellular domain that also shows homology to EGFR. Its
CC overexpression correlates with a poor prognosis in breast and
CC ovarian cancers. The invention provides Her-2/neu fusion
CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its DeltaEP fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.
XX
SQ Sequence 1255 AA;
Query Match 97.4%; Score 6602; DB 23; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;
Qy 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLPASPETHLDMRLHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLPASPETHLDMRLHLYQGCVVQGNL 60
Qy 61 ELTYLPTNASLSFLODIOEVQGYVLIHQNQVQLRLRIVRGTLQDFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLODIOEVQGYVLIHQNQVQLRLRIVRGTLQDFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILMKDIFHKNNQLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILMKDIFHKNNQLA 180
Qy 181 LTLIDNTRSRACHPCSPCKRCWGSESDCOSLTRTVCAAGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDNTRSRACHPCSPCKRCWGSESDCOSLTRTVCAAGCARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
-Y 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 ELEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFTELEETIGLYLISAWPDSL 420
Db 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFTELEETIGLYLISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGSIWGLSLRELGSGLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGSIWGLSLRELGSGLALIHNNTHLCFVHTV 480
Qy 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCNCVSFLRGQEC 540
Db 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCNCVSFLRGQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECCQPNQSVTCFGEADQCVCAHAKYDPPFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECCQPNQSVTCFGEADQCVCAHAKYDPPFCVARC 600
Qy 601 PSGVKPDLSYMPIWKFPPDEEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTISVNAVVG 660
Db 601 PSGVKPDLSYMPIWKFPPDEEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTISVNAVVG 660
Qy 661 ILLVVLGVVFGILII---QYIKANS--KFIGITEL--PLTPSGAMPNQAQRILKETEL 712
Db 661 ILLVVLGVVFGILIIKRRQOKIRKYMRRLLOETELVELPLTPSGAMPNQAQRILKETEL 720
Qy 713 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 772
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780

QY 773 YVSRLLGICLTSTVQLVQVLTQMPYGCCLLDHVRNRRGRGSLQDNLNWCQIAKMSYLEDVR 832
DB 781 YVSRLLGICLTSTVQLVQVLTQMPYGCCLLDHVRNRRGRGSLQDNLNWCQIAKMSYLEDVR 840
QY 833 LVHRDLAARNVLKSPNHVKTIDFGLARLLDIDETEHADGGKVPKMALESILRRFT 892
DB 841 LVHRDLAARNVLKSPNHVKTIDFGLARLLDIDETEHADGGKVPKMALESILRRFT 900
QY 893 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWCWM 952
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWCWM 960
QY 953 IDSECRPRFRELSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1012
DB 961 IDSECRPRFRELSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1020
QY 1013 EYLVPQOGEFCDDPAPAGAGWVHRRSSSTRSGGDLTLCGLPSEEBAPRSPAPSEG 1072
DB 1021 EYLVPQOGEFCDDPAPAGAGWVHRRSSSTRSGGDLTLCGLPSEEBAPRSPAPSEG 1080
QY 1073 AGSDVFDGDLGMGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1132
DB 1081 AGSDVFDGDLGMGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1133 NQDVRPQPPSPREGPLPAAPAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1192
DB 1141 NQDVRPQPPSPREGPLPAAPAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
QY 1193 GGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGDLVDPV 1247
DB 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGDLVDPV 1255

RESULT 14

AAU77114

ID AAU77114 standard; Protein; 1255 AA.

AC AAU77114;

XX

DT 05-JUN-2002 (first entry)

XX

XX Human Her-2/neu polypeptide.

DE

XX

KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.

XX

OS Homo sapiens.

XX

XX WO200213847-A2.

XX

XX 21-FEB-2002.

XX

XX 13-AUG-2001; 2001WO-US25408.

XX

XX 14-AUG-2000; 2000US-0638280.

XX

XX 28-SEP-2000; 2000US-0675904.

XX

XX (CORI-) CORIXA CORP.

PA

XX

XX Gaiger A, Cheever MA, Hand-zimmermann S;

XX

XX WPI; 2002-280741/32.

XX

XX N-PSDB; ABK10730.

XX

XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide

XX

PS Disclosure; Page 71-74; 74pp; English.

XX

XX

XX

XX

XX

XX

XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myeloma, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX

SQ Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 23; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;

QY 1 MELAALCRWGLLLALLPPGAASQTCTGTDMKRLPASPETHLDMRLHLYGCGVQVQNL 60
DB 1 MELAALCRWGLLLALLPPGAASQTCTGTDMKRLPASPETHLDMRLHLYGCGVQVQNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTLFEDNYALAVLNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTLFEDNYALAVLNG 120
QY 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNOLA 180
DB 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNOLA 180
QY 181 LTLIDTNRSRACHPCSPMKGSRGWCSESDCQSLTRTVACAGGCARCKPLTDCCHQC 240
DB 181 LTLIDTNRSRACHPCSPMKGSRGWCSESDCQSLTRTVACAGGCARCKPLTDCCHQC 240
QY 241 AAGCTGPKHSDCLACLPNHSIGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLPNHSIGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVCPLNQEVTAEDGTQCEKSKPCARVCYGLGMQYIKANSKFIGIT 360
DB 301 YNYLSTDVSGCTLVCPLNQEVTAEDGTQCEKSKPCARVCYGLGMQYIKANSKFIGIT 360
QY 361 ELSEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLEETIGLYLISAWPDSL 420
DB 361 IQSEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLEETIGLYLISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRLSLRELGLSLALHNNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRLSLRELGLSLALHNNTHLCFVHTV 480
QY 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHOLCARGHCWGPCTOCVNCQOFILRGQEC 540
DB 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHOLCARGHCWGPCTOCVNCQOFILRGQEC 540
QY 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
DB 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCPCQASPLTSTISAVVG 660
DB 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCPCQASPLTSTISAVVG 660
QY 661 ILLVVLGVVFGILI---OYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETTEL 712
DB 661 ILLVVLGVVFGILI---OYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETTEL 712
QY 713 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVNAVGVSP 772
DB 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVNAVGVSP 780
QY 773 YVSRLLGICLTSTVQLVQVLTQMPYGCCLLDHVRNRRGRGSLQDNLNWCQIAKMSYLEDVR 832
DB 781 YVSRLLGICLTSTVQLVQVLTQMPYGCCLLDHVRNRRGRGSLQDNLNWCQIAKMSYLEDVR 840
QY 833 LVHRDLAARNVLKSPNHVKTIDFGLARLLDIDETEHADGGKVPKMALESILRRFT 892

Db 841 LVHRDLARNVVKSPNHVKITDFGLARLLDIDETEHADGCKVPIKMALESILRRFT 900
Qy 893 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 952
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
Qy 953 IDSECRPRFELVSEFSRMDRPFVVIQNEDELGPASPLDSTFYRSLLDDMDGLVDA 1012
Db 961 IDSECRPRFELVSEFSRMDRPFVVIQNEDELGPASPLDSTFYRSLLDDMDGLVDA 1020
Qy 1013 EYLVLPQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1072
Db 1021 EYLVLPQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Qy 1073 AGSDVFDGDLGMAAGKGLSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQBEYV 1132
Db 1081 AGSDVFDGDLGMAAGKGLSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQBEYV 1140
Qy 1133 NOPDVRPOPSPREGPLPAARPAGATLERAKTLPNGKGVVVDVFAFGGAVENPEYLTPO 1192
Db 1141 NOPDVRPOPSPREGPLPAARPAGATLERPKTLPNGKGVVVDVFAFGGAVENPEYLTPO 1200
Qy 1193 GGAAPQHPHPPAFSPADNLVYWDQDPERGAPPSTFKGTPTABNPEYLGIDVPV 1247
Db 1201 GGAAPQHPHPPAFSPADNLVYWDQDPERGAPPSTFKGTPTABNPEYLGIDVPV 1255

RESULT 15

AAR39568
ID AAR39568 standard; Protein; 1433 AA.

XX AAR39568;

DT 07-FEB-1994 (first entry)

DE Sequence of c-erbB-2 tumour antigen.

KW Tumour antigen; c-erbB-2; glycoprotein.

XX Homo sapiens.

XX WO9316185-A.

XX 19-AUG-1993.

XX 05-FEB-1993; 93WO-US01055.

XX 06-FEB-1992; 92US-0831967.

XX (CETU) CETUS ONCOLOGY CORP.
XX (CREA-) CREATIVE BIOMOLECULES INC.

XX Houston LL, Huston JS, Oppermann H, Ring DB;

XX WPI; 1993-272889/34.

XX N-PSDB; AAQ46083.

XX New single chain Fv polypeptide binding to C-erbB-2 tumour

XX antigen - for imaging or treating breast or ovarian cancer etc.

XX Disclosure; pages 48-54; 87pp; English.

XX c-erbB-2 refers to a protein antigen expressed on the surface of
XX tumour cells, such as breast and ovarian tumour cells, which is an
XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
XX the location of a stop codon in AAQ46083.

XX Sequence 1433 AA;

Query Match 96.8%; Score 6559; DB 14; Length 1433;
Best Local Similarity 96.7%; Pred. No. 0;

Matches 1213; Conservative 11; Mismatches 23; Indels 8; Gaps 3;
Qy 1 MELAALCRWGLLLALLPFGAASCTGCTDMKRLPASPETHLDMKRLHYGCCVQVQGNL 60
Db 1 MELAALCRWGLLLALLPFGAASCTGCTDMKRLPASPETHLDMKRLHYGCCVQVQGNL 60
Qy 61 ELTYLPTNASISFLQDIOEVQYVLIIAHNOVQVPLQRLRIVRGTPOLFEDNVALAVLNG 120
Db 61 ELTYLPTNASISFLQDIOEVQYVLIIAHNOVQVPLQRLRIVRGTPOLFEDNVALAVLNG 120
Qy 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGVLIQORNPOLCYQDITLWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGVLIQORNPOLCYQDITLWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLLVPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLLVPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360
Qy 361 ELBPAGCKITFGSLAFIPESPFGDPASNTAPLQPEQLQVFPETLBEITCYLYISAWPDSL 420
Db 361 IQEPAGCKITFGSLAFIPESPFGDPASNTAPLQPEHLQVFPETLEQITCYLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSITLQGLISWLGRLSRLGSLGALIHNNTHLSFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSITLQGLISWLGRLSRLGSLGALIHNNTHLSFVHTV 480
Qy 481 PWDQLFRNPHOALLHTANRPEDECVEGLACHQLCARGHGWGPGTQCNCVSQFLRGQBC 540
Db 481 PWDQLFRNPHOALLHTANRPEDECVEGLACHQLCARGHGWGPGTQCNCVSQFLRGQBC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
Qy 601 PSGVKPDLSTYMPIWKFPDEEGACOPCPINCTHSCVDLDDKGCAPABORASPLTSIVSAVVG 660
Db 601 PSGVKPDLSTYMPIWKFPDEEGACOPCPINCTHSCVDLDDKGCAPABORASPLTSIVSAVVG 660
Qy 661 ILLVVVLGVVFGILII---QYIKANS--KFICITEL--PLTPSGAMPNQAOHRIILKETEL 712
Db 661 ILLVVVLGVVFGILIIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAOHRIILKETEL 720
Qy 713 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVQSP 772
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVQSP 780
Qy 773 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRENRGLSGDLLNWCMIKAGMSYLEVDV 832
Db 781 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRENRGLSGDLLNWCMIKAGMSYLEVDV 840
Qy 833 LVHRDLARNVVKSPNHVKITDFGLARLLDIDETEHADGCKVPIKMALESILRRFT 892
Db 841 LVHRDLARNVVKSPNHVKITDFGLARLLDIDETEHADGCKVPIKMALESILRRFT 900
Qy 893 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 952
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
Qy 953 IDSECRPRFELVSEFSRMDRPFVVIQNEDELGPASPLDSTFYRSLLDDMDGLVDA 1012
Db 961 IDSECRPRFELVSEFSRMDRPFVVIQNEDELGPASPLDSTFYRSLLDDMDGLVDA 1020
Qy 1013 EYLVLPQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1072
Db 1021 EYLVLPQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080

Qy	1073	AGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEYV	1132
Db	1081	AGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1133	NOPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAPGGAIVENPEYLTTPQ	1192
Db	1141	NOPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAPGGAIVENPEYLTTPQ	1200
Qy	1193	GGAAPQPHPPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV	1247
Db	1201	GGAAPQPHPPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 09:16:33
Job time : 41.6984 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-579-593-12

Perfect score: 6803

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPAENPEYLGLDVVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6657	97.9	1255	21	Human heregulin 2
2	6657	97.9	1255	22	Human tyrosine kin
3	6657	97.9	1255	22	HER2 transgene pla
4	6657	97.9	1255	23	Human HER2 (ErbB2)
5	6651	97.8	1255	17	HER-2/neu protein.
6	6651	97.8	1255	20	Human HER-2/neu on
7	6651	97.8	1255	21	Human HER-2/neu pr
8	6651	97.8	1255	21	Amino acid sequenc
9	6651	97.8	1255	22	Human HER-2/neu pr
10	6651	97.8	1255	22	HER2/neu amino aci

11	6651	97.8	1255	23	AAE24067	Human Her-2 protei
12	6651	97.8	1255	23	AAE20479	Human Her-2/neu pr
13	6651	97.8	1255	23	AAE20479	Human Her-2/neu pr
14	6651	97.8	1255	23	AAU77114	Human Her-2/neu po
15	6651	97.8	1255	23	AAU77114	Human Her-2/neu po
16	6651	97.8	1255	23	AAU77114	Sequence of c-erbB
17	6651	97.8	1255	23	AAU77114	Human breast cance
18	6651	97.8	1255	23	AAU77114	Human breast cance
19	6651	97.8	1255	23	AAU77114	Human breast cance
20	6651	97.8	1255	23	AAU77114	Human breast cance
21	6651	97.8	1255	23	AAU77114	Human breast cance
22	6651	97.8	1255	23	AAU77114	Human breast cance
23	6651	97.8	1255	23	AAU77114	Human breast cance
24	6651	97.8	1255	23	AAU77114	Human breast cance
25	6651	97.8	1255	23	AAU77114	Human breast cance
26	6651	97.8	1255	23	AAU77114	Human breast cance
27	6651	97.8	1255	23	AAU77114	Human breast cance
28	6651	97.8	1255	23	AAU77114	Human breast cance
29	6651	97.8	1255	23	AAU77114	Human breast cance
30	6651	97.8	1255	23	AAU77114	Human breast cance
31	6651	97.8	1255	23	AAU77114	Human breast cance
32	6651	97.8	1255	23	AAU77114	Human breast cance
33	6651	97.8	1255	23	AAU77114	Human breast cance
34	6651	97.8	1255	23	AAU77114	Human breast cance
35	6651	97.8	1255	23	AAU77114	Human breast cance
36	6651	97.8	1255	23	AAU77114	Human breast cance
37	6651	97.8	1255	23	AAU77114	Human breast cance
38	6651	97.8	1255	23	AAU77114	Human breast cance
39	6651	97.8	1255	23	AAU77114	Human breast cance
40	6651	97.8	1255	23	AAU77114	Human breast cance
41	6651	97.8	1255	23	AAU77114	Human breast cance
42	6651	97.8	1255	23	AAU77114	Human breast cance
43	6651	97.8	1255	23	AAU77114	Human breast cance
44	6651	97.8	1255	23	AAU77114	Human breast cance
45	6651	97.8	1255	23	AAU77114	Human breast cance

ALIGNMENTS

RESULT 1
ID AAY92620 standard; Protein; 1255 AA.
AC AAY92620;
DT 10-AUG-2000 (first entry)
XX Human heregulin 2 (Her2).
DE Human heregulin 2 (Her2).
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Domain 1..173
FT /label= N-terminal
FT /note= "mature polypeptide"
FT Region 5..25
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 59..73
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 103..117
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 149..163
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Domain 174..323

FT Region /label= Cysteine_rich_domain
 FT 210..224 /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 250..264
 FT Region /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 324..483
 FT Domain /label= Ligand_binding_domain
 FT 325..339
 FT Region /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 369..383
 FT Region /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 465..479
 FT Region /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 484..623
 FT Domain /label= Cysteine_rich_domain
 FT 579..593
 FT Region /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 624..654
 FT Domain /label= Transmembrane_domain
 FT 632..652
 FT Region /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 653..667
 FT Region /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 655..1010
 FT Domain /label= Tyrosine_kinase_domain
 FT 651..675
 FT Region /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 695..709
 FT Region /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 710..730
 FT Region /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 1011..1235
 FT Domain /label= C-terminal_domain
 FT
 XX WO200020027-A2.
 XX 13-APR-2000.
 XX 05-OCT-1999; 99WO-DK00525.
 XX 05-OCT-1998; 98DK-0001261.
 PR 20-OCT-1998; 98US-0105011.
 XX (MEBI-) M & E BIOTECH AS.
 XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 PI Gautam A, Birk P, Karlsson G;
 PI WPI; 2000-349917/30.
 DR N-PSDB; AAA09455.
 XX Inducing immune responses to weakly immunogenic, tumor associated
 PT peptide antigens for the treatment of breast and prostate cancer
 XX Claim 62; Page 193-198; 220pp; English.
 PS This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Subdominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 XX

Sequence 1255 AA;

Query Match 97.9%; Score 6657; DB 21; Length 1255;
 Best Local Similarity 97.8%; Pred. No. 0;
 Matches 1227; Conservative 9; Mismatches 19; Indels 0; Gaps 0;
 QY 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLPASPETHLDMLRHLYQGCVVQGNL 60
 DB 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLPASPETHLDMLRHLYQGCVVQGNL 60
 QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNOVQVPLQRLIRIVRGTOLEFEDNALAVLDNG 120
 DB 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNOVQVPLQRLIRIVRGTOLEFEDNALAVLDNG 120
 QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKKNOLA 180
 DB 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKKNOLA 180
 QY 181 LTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240
 DB 181 LTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQRCCKSPCARVCYGLGMQVIKANSKFIGIT 360
 DB 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQRCCKSPCARVCYGLGMHLREVRVTSAN 360
 QY 361 ELEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFETLEEITGYLYISAWPDSL 420
 DB 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFETLEEITGYLYISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
 QY 481 PWDQLFRNPQALHTANRPEDECVGEGLAHQLCARGHCWPGPTQCNCVSFLRGQEC 540
 DB 481 PWDQLFRNPQALHTANRPEDECVGEGLAHQLCARGHCWPGPTQCNCVSFLRGQEC 540
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVCACAHYKDPFPCVARC 600
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVCACAHYKDPFPCVARC 600
 QY 601 PQYIKANSKFIGITELPDEGACQPCPINCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660
 DB 601 PSGVXPDSLUSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660
 QY 661 ILLVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQRILKETEL 720
 DB 661 ILLVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQRILKETEL 720
 QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGS 780
 DB 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGS 780


```
Qy 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
Dy 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTPQMPYGCCLDHVRENRLGSGDQLLNCWQIAKGMVLEDR 840
Dy 781 YVSRLLGICLTSTVQLVTPQMPYGCCLDHVRENRLGSGDQLLNCWQIAKGMVLEDR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDGLARLLDDIETEHADGKVPKIMWALESTILRRFT 900
Dy 841 LVHRDLAARNVLKSPNHVKITDGLARLLDDIETEHADGKVPKIMWALESTILRRFT 900
Qy 901 HQSDWMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
Dy 901 HQSDWMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
Qy 961 IDSECRPRFRELVEFSFMRDPPORFVVIQNEIDLPASPLDSTFYRSLLDDMGDLVDA 1020
Dy 961 IDSECRPRFRELVEFSFMRDPPORFVVIQNEIDLPASPLDSTFYRSLLDDMGDLVDA 1020
Qy 1021 EYILVPOQGFPCPDAPAGAGGMVHRHRSSTRSGGDLTLGLPSEBEEAPRSLAPSEG 1080
Dy 1021 EYILVPOQGFPCPDAPAGAGGMVHRHRSSTRSGGDLTLGLPSEBEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGKGLQSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Dy 1081 AGSDVFDGDLGMAAGKGLQSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPGATLERAKTLSPGKNGVVDVFAFGAVENPEYLTPO 1200
Dy 1141 NOPDVRPQPPSPREGPLPAARPGATLERAKTLSPGKNGVVDVFAFGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPFDNLVYWDQDPPERCAPPSTFKGTPTAENPEYLGIDVPV 1255
Dy 1201 GGAAPQHPHPPAFSPFDNLVYWDQDPPERCAPPSTFKGTPTAENPEYLGIDVPV 1255
```

RESULT 3

```
ID AAB60167 standard; Protein; 1255 AA.
XX
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
DE HER2 transgene plasmid construct encoded protein.
```

KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer; antibody.

XX Homo sapiens.

OS Synthetic.

XX W0200100244-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-US17229.

XX 25-JUN-1999; 99US-0141316.

XX 16-MAR-2000; 2000US-0189844.

XX (GETH) GENENTECH INC.

XX Erickson S, Schwall R;

XX WPI; 2001-061962/07.

XX N-PSDB; AAF24297.

XX Treating tumors, particularly breast cancers, which overexpress an ErbB receptor and does not respond to an anti-ErbB antibody, comprises conjugating the antibody to a maytansinoid -

Example 3; Fig 4; 92pp; English.

CC The present invention provides a method of treating cancer by administering a conjugate of anti-ErbB antibody with a maytansinoid. In particular, the antibody is directed against ErbB2 (also known as HER2 and p185neu). The method is particularly useful in the treatment of breast, ovarian, stomach, endometrial, salivary gland, lung, kidney, colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

XX Sequence 1255 AA;

Qy Query Match 97.9%; Score 6657; DB 22; Length 1255;
Dy Best Local Similarity 97.8%; Pred. No. 0;
Matches 1227; Conservative 9; Mismatches 19; Indels 0; Gaps 0;

```
Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKLRLPASPETHLDMRLHYGQCQVQGNL 60
Dy 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKLRLPASPETHLDMRLHYGQCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVGYVLIAHNVQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Dy 61 ELTYLPTNASLSFLQDIQEVGYVLIAHNVQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Qy 121 DPLNNTTPTVGTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNOLA 180
Dy 121 DPLNNTTPTVGTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCCAGGCARCKGPLPTDCCHQEC 240
Dy 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCCAGGCARCKGPLPTDCCHQEC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGYTTFGASCVTACP 300
Dy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGYTTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHNOEVTADGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
Dy 301 YNYLSTDVGSCTLVCPLNHNOEVTADGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFLPESPDGPASNTAPLQPEQLQVFTLEBITGYLIASWPDSL 420
Dy 361 IQEFAGCKKIFGSLAFLPESPDGPASNTAPLQPEQLQVFTLEBITGYLIASWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLSLQGLISWLGRLSRLRELGSLALHNNHLCFVHTV 480
Dy 421 DLSVFQNLQVIRGRILHNGAYSLSLQGLISWLGRLSRLRELGSLALHNNHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Dy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGPEADQCVACAHYKDPFPCVARC 600
Dy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGPEADQCVACAHYKDPFPCVARC 600
Qy 601 POYIKANSKFIGITELPDEEGACOPCPINCTHSCVDLDDKGCPCAEORASPLTSIVSAVVG 660
Dy 601 PSYGKPDLSYWPWKPFDEEGACOPCPINCTHSCVDLDDKGCPCAEORASPLTSIVSAVVG 660
Qy 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKETEL 720
Dy 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
Dy 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTPQMPYGCCLDHVRENRLGSGDQLLNCWQIAKGMVLEDR 840
Dy 781 YVSRLLGICLTSTVQLVTPQMPYGCCLDHVRENRLGSGDQLLNCWQIAKGMVLEDR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDGLARLLDDIETEHADGKVPKIMWALESTILRRFT 900
Dy 841 LVHRDLAARNVLKSPNHVKITDGLARLLDDIETEHADGKVPKIMWALESTILRRFT 900
```

Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRFT 900
 Qy 901 HQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVTYIMVWKWM 960
 Db 901 HQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVTYIMVWKWM 960
 Qy 961 IDSECRPFRELVSFSESRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDGDLVDA 1020
 Db 961 IDSECRPFRELVSFSESRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDGDLVDA 1020
 Qy 1021 EYLVPQOGFFCPDPAPGAGGVMHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 Db 1021 EYLVPQOGFFCPDPAPGAGGVMHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 Qy 1081 AGSVDGDLGMAAGKLOSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPYV 1140
 Db 1081 AGSVDGDLGMAAGKLOSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPYV 1140
 Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVDVFAFGAVENPEYLTPQ 1200
 Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVDVFAFGAVENPEYLTPQ 1200
 Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPPTAENPEYLGLDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPPTAENPEYLGLDVPV 1255

RESULT 4

AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 XX
 AC AAU74545;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Human HER2 (ErbB2) polypeptide.
 XX
 KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glial disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macrophagal disorder; epithelial disorder;
 KW stromal disorder; blastocoealic disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.
 XX
 OS Homo sapiens.
 XX
 PN US2002001587-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 16-MAR-2001; 2001US-0811123.
 XX
 PR 16-MAR-2000; 2000US-189844P.
 PR 05-OCT-2000; 2000US-238327P.
 XX
 PA (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIW/) SLIWKOWSKI M.
 XX
 PI Erickson S, Schwall R, Sliwkowski M;
 XX
 DR WPI; 2002-163686/21.
 DR N-PSDB; ABK14058.
 XX
 PT Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor, ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
 XX
 PS Example 3; Fig 7; 93pp; English.
 PS
 XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.
 XX

SQ Sequence 1255 AA;

Query Match 97.9%; Score 6657; DB 23; Length 1255;
 Best Local Similarity 97.8%; Pred. No. 0;
 Matches 1227; Conservative 9; Mismatches 19; Indels 0; Gaps 0;
 Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRPASPETHLDMRLHLYGCGVQVGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRPASPETHLDMRLHLYGCGVQVGNL 60
 Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIHQNVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120
 Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIHQNVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120
 Qy 121 DPLNNTTPVTGASPGGLREQLRLSITELKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
 Db 121 DPLNNTTPVTGASPGGLREQLRLSITELKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
 Qy 181 LTLIDTNRSRACHPCSPMCKSGRSGESSEDCQSILTRTVACGGCARGKPLTDCCHEOC 240
 Db 181 LTLIDTNRSRACHPCSPMCKSGRSGESSEDCQSILTRTVACGGCARGKPLTDCCHEOC 240
 Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPHLNQVTAEDGTQRCCKSKPCARVCYGLGMOVIKANSKFIGIT 360
 Db 301 YNYLSTDVGSCTLVCPHLNQVTAEDGTQRCCKSKPCARVCYGLGMOVIKANSKFIGIT 360
 Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFTLEITCYLYISAMPDLSL 420
 Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFTLEITCYLYISAMPDLSL 420
 Qy 421 DLSVFQNLQVIRGIRLHNGAYSLTLQGLIGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGIRLHNGAYSLTLQGLIGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480
 Qy 481 PWDOLFNRPHOALLHTANRPEDECVGEGGLACHQICARGHCWGPGPTQCVNCSQFLRGQEC 540
 Db 481 PWDOLFNRPHOALLHTANRPEDECVGEGGLACHQICARGHCWGPGPTQCVNCSQFLRGQEC 540
 Qy 541 VEECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
 Db 541 VEECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
 Qy 601 PQVIKANSKFIGITELPDEEGACQPCPINCTHSCVDLDDKGCAPAEORASPLTSTVSAVVG 660
 Db 601 PSGVKPDLSYMPITWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEORASPLTSTVSAVVG 660
 Qy 661 ILLVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 Qy 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETILDEAYVNAVGVSP 780
 Db 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETILDEAYVNAVGVSP 780
 Qy 781 YVSRLLGICLTSTVQLVTQMLPMPYQCLLDHVRNRRGLGSQDLLNWCNQIAGKHSYLEDVR 840

Db 781 YVSRLLGICLTSTVQLVTPQMPYGCILLDHVRENRRGLSQDILLNWCMIKAGMSYLEDVR 840
QY LVHRDLAARNVLVAKPNHVKITDFGLARLLDIDETEHADGCKVPIKMALESILRRRT 900
Db 841 LVHRDLAARNVLVAKPNHVKITDFGLARLLDIDETEHADGCKVPIKMALESILRRRT 900
QY HQSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVTYMIWVKWM 960
Db 901 HQSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVTYMIWVKWM 960
QY 961 IDSECRPRELVSEFSRMRDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
Db 961 IDSECRPRELVSEFSRMRDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
QY 1021 EYLVVPOQGFCCPDAPAGAGGVMVHRHRSSTRSGGDLTLGLEPSEERAPRSLAPSEG 1080
Db 1021 EYLVVPOQGFCCPDAPAGAGGVMVHRHRSSTRSGGDLTLGLEPSEERAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQPDVVRPQPPSPREGPLPAARAGATLERAKTSLPGKNGVVKDVFAGGAVENPEYLTPO 1200
Db 1141 NQPDVVRPQPPSPREGPLPAARAGATLERAKTSLPGKNGVVKDVFAGGAVENPEYLTPO 1200
QY 1201 GGAAQPHPPPPAFSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAQPHPPPPAFSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 5

AAW01111

ID AAW01111 standard; Protein; 1255 AA.

XX AC AAW01111;

XX 01-JAN-1997 (first entry)

XX HER-2/neu protein.

XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;

XX breast cancer; ovary cancer; colon cancer; lung cancer;

XX prostate cancer; immunisation; tumour; vaccine; vector.

XX OS Homo sapiens.

XX FT Key Location/Qualifiers

XX FT Domain 676..1255

XX FT /label= Intracellular domain

XX FT /note= "claimed domain, useful for immunisation"

XX PN WO9630514-Al.

XX PD 03-OCT-1996.

XX PF 28-MAR-1996; 96WO-US01689.

XX PR 31-MAR-1995; 95US-0414417.

XX PA (UNITW) UNIV WASHINGTON.

XX PI Cheever MA, Disis ML;

XX DR WPI; 1996-455361/45.

XX DR N-PSDB; AAT40739.

XX PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or

XX PT treatment of malignancies with which the HER-2/neu oncogene is

XX PT associated

XX PS Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.

XX Sequence 1255 AA;

Query Match 97.8%; Score 6651; DB 17; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASQTCTGTDMLRLPASPETHLDMLRHLYQGCQVVGNNL 60

Db 1 MELAALCRWGLLLALLPPGAASQTCTGTDMLRLPASPETHLDMLRHLYQGCQVVGNNL 60

QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVROVPLQRLRIVRGTQLPEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVROVPLQRLRIVRGTQLPEDNYALAVLDNG 120

QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180

Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180

QY 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACAGCARCKGPLPTDCCHEQC 240

Db 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACAGCARCKGPLPTDCCHEQC 240

QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300

QY 301 NYNLSTDVSGSCTLVCPHLNQEVTAEDGTQRCBCKSPCARVCVGLGMQVIKANSKFIT 360

Db 301 NYNLSTDVSGSCTLVCPHLNQEVTAEDGTQRCBCKSPCARVCVGLGMQHLRVRVTSAN 360

QY 361 ELEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFETLEEITGYLYISAWPDSLP 420

Db 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFETLEEITGYLYISAWPDSLP 420

QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGURSRELGLALIHNNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGURSRELGLALIHNNTHLCFVHTV 480

QY 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540

Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540

QY 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600

Db 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600

QY 601 POYIKANSKFIGITELPDEEGACQPCINCHTSCVDLDDKGPAPORASPLTSISAVVG 660

Db 601 PSQVPRDLSYMPDWKPFDEEGACQPCINCHTSCVDLDDKGPAPORASPLTSISAVVG 660

QY 661 ILLVVVLGVVFGILIKRRQKIRKYMRLRLQETELVEPLTPSGAMPNOAQRILKETEL 720

Db 661 ILLVVVLGVVFGILIKRRQKIRKYMRLRLQETELVEPLTPSGAMPNOAQRILKETEL 720

QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780

Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780

QY 781 YVSRLLGICLTSTVQLVTPQMPYGCILLDHVRENRRGLSQDILLNWCMIKAGMSYLEDVR 840

Db 781 YVSRLLGICLTSTVQLVTPQMPYGCILLDHVRENRRGLSQDILLNWCMIKAGMSYLEDVR 840

```
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALESILRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALESILRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVVKWM 960
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVVKWM 960
QY 961 IDSECRFRRELVSERFARMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDGDLVDA 1020
DB 961 IDSECRFRRELVSERFARMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDGDLVDA 1020
QY 1021 EYLVPOQGFPCPDPAAGAGGVHHRSSSTRSGGGDLTLGLPSEBEEAPRSLAPSEG 1080
DB 1021 EYLVPOQGFPCPDPAAGAGGVHHRSSSTRSGGGDLTLGLPSEBEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGKGLQSLPTHDPSPQLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAAGKGLQSLPTHDPSPQLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NOPDVRQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFATGGAVENPEYLTPO 1200
DB 1141 NOPDVRQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFATGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
DB 1201 GGAAPQHPPPAPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 6
ID AAW92406
XX AAW92406 standard; Protein; 1255 AA.
AC AAW92406;
XX
XX
DT 21-APR-1999 (first entry)
XX
DE Human HER-2/neu oncogene protein.
XX
KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW malignancy; treatment; tumour.
XX
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX Region 676..1255
XX /note= "region which elicits immune response"
XX
XX US5869445-A.
XX
XX 09-FEB-1999.
XX
XX 01-APR-1996; 96US-0625101.
XX
XX 01-APR-1996; 96US-0625101.
XX 17-MAR-1993; 93US-0033644.
XX 12-AUG-1993; 93US-0106112.
XX 31-MAR-1995; 95US-0414417.
XX
XX (UNIW ) UNIV WASHINGTON.
XX
XX Cheever MA, Disis ML;
XX
XX WPI; 1999-152835/13.
XX N-PSDB; AAX01912.
XX
XX Use of HER-2/neu polypeptides - for eliciting an immune response to
XX an HER-2/neu associated malignancy, particularly for treating or
XX preventing tumours
XX
XX Claim 3; Column 31-38; 26pp; English.
XX
XX This sequence represents the human HER-2/neu oncogene protein. A fragment
```

```
CC of this protein is used in a method for eliciting or enhancing an immune
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.
XX
SQ Sequence 1255 AA;
Query Match 97.8%; Score 6651; DB 20; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1255; Conservative 10; Mismatches 20; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMRLHLYGQCQVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMRLHLYGQCQVQGNL 60
C/
61 ELTVLPTNASLSFLQDIOEVQGVYLIQHNQVQVPLQRLIRVGTQLPEDNYALAVLNG 120
DB 61 ELTVLPTNASLSFLQDIOEVQGVYLIQHNQVQVPLQRLIRVGTQLPEDNYALAVLNG 120
QY 121 DPLNNTTPTVGASPGGLREQLRLSTELILKGVLIQRPOLCYQDTILWKDILFHNKOLA 180
DB 121 DPLNNTTPTVGASPGGLREQLRLSTELILKGVLIQRPOLCYQDTILWKDILFHNKOLA 180
QY 181 LTLIDTNRSRACHPCSPMCKGSCWGESSEDCQSLTRTVCAAGCARCKGPLTDCCHQC 240
DB 181 LTLIDTNRSRACHPCSPMCKGSCWGESSEDCQSLTRTVCAAGCARCKGPLTDCCHQC 240
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSTLVCPPLHNOEVTAEQDGTQCEKSKPCARVCYGLHMOYIKANSKFIGIT 360
DB 301 YNYLSTDVGSTLVCPPLHNOEVTAEQDGTQCEKSKPCARVCYGLHMOYIKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAFPLESFDGDPASNTAPLQPEQLQVFTLEETITGLYISAMPDLSL 420
DB 361 IQEFAGCKKIFGSLAFPLESFDGDPASNTAPLQPEQLQVFTLEETITGLYISAMPDLSL 420
QY 421 DLSVFQNLQVIRGRIHLHNGAYSILTLQGLGISWLGSLRLSRELGSGLALIHNTLHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRIHLHNGAYSILTLQGLGISWLGSLRLSRELGSGLALIHNTLHLCFVHTV 480
QY 481 PWQLPRNPHOALLHTANRPEDECVGEGLAACHOL-CARGHCGPGTQCVNCSQFLRGQEC 540
DB 481 PWQLPRNPHOALLHTANRPEDECVGEGLAACHOL-CARGHCGPGTQCVNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 601 PQYIKANSKFIGITELPDEBACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTVSAVVG 660
DB 601 PSQVXPDLSTYMPIWKFPDEBACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKVTMRLLQETELVEPLTPSCAMPNQAMRILKETEL 720
DB 661 ILLVVVLGVVFGILIKRQOKIRKVTMRLLQETELVEPLTPSCAMPNQAMRILKETEL 720
QY 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKEITLDEAYVMAGVGP 780
DB 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKEITLDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVTQMPYVGLDHDHVRNRLGSDLLNWCNQIAKGSYLESDVR 840
DB 781 YVSRLLGICLTSTVQLVTQMPYVGLDHDHVRNRLGSDLLNWCNQIAKGSYLESDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALESILRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALESILRRFT 900
```

QY 901 HQSDVMSYCVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVTMIMVKWM 960
DB HQSDVMSYCVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVTMIMVKWM 960
QY 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
DB IDSECRPRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
QY 1021 EYLVPQOQFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPBSPLAPSEG 1080
DB EYLVPQOQFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPBSPLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTPLPSETDGYVAPLTCSPQPEYV 1140
DB AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NOPDVRPQPPSPREGPLPAARAGATLERAKTSLSPKNGVVKDVFAGGAVENPEYLTPO 1200
DB NOPDVRPQPPSPREGPLPAARAGATLERAKTSLSPKNGVVKDVFAGGAVENPEYLTPO 1200
QY 1201 GGAAPQPHPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
DB GGAAPQPHPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 7

AAB21198
ID AAB21198 standard; protein; 1255 AA.
XX AAB21198;
AC AAB21198;
XX 12-JAN-2001 (first entry)
DE Human HER-2/neu protein.
XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.
XX Homo sapiens.
XX WO200044899-A1.
XX 03-AUG-2000.
XX 28-JAN-2000; 2000WO-US02164.
XX 29-JAN-1999; 99US-0117976.
XX (CORI-) CORIXA CORP.
XX (SMIK) SMITHKLINE BEECHAM.
XX Cheever MA, Gheyene D;
XX WPI; 2000-505976/45.
XX N-PSDB; AAA89736.
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX Claim 52; Fig 7; 128pp; English.
XX The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
XX
SQ Sequence 1255 AA;
Query Match 97.8%; Score 6651; DB 21; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPGAASTQVCTGDMKRLPASPETHLDMLRHLYQCCOVQGNL 60
DB MELAALCRWGLLLALLPGAASTQVCTGDMKRLPASPETHLDMLRHLYQCCOVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLIRVRGTQLFEDNYALVDNG 120
DB ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLIRVRGTQLFEDNYALVDNG 120
QY 121 DPLNTTPTVTCASPGGLRELQRLSUTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
DB DPLNTTPTVTCASPGGLRELQRLSUTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCHEOC 240
DB LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMQVYIKANSKFIGIT 360
DB YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMQVYIKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFEITLGYLYISAMPDLSLP 420
DB IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFEITLGYLYISAMPDLSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISHWGLRSLRELGSGLALIHNNTHLCFVHTV 480
DB DLSVFQNLQVIRGRILHNGAYSLTLQGLGISHWGLRSLRELGSGLALIHNNTHLCFVHTV 480
QY 481 PWDQLFRNPHOALLHTANRDECVGEGGLACHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
DB PWDQLFRNPHOALLHTANRDECVGEGGLACHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
QY 541 VEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPPECVARC 600
DB VEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPPECVARC 600
QY 601 POYIKANSKFIGITELPDEEGACQPCINCTHSCVDLDDKGCAPAQASPLTSIVSAVVG 660
DB PSQVKPDLSTYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPAQASPLTSIVSAVVG 660
QY 661 ILLVVVLGVTFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQRILKETEL 720
DB ILLVVVLGVTFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQRILKETEL 720
QY 721 RKVKVLGGAGFTVYKGIWIPDGENVKIPVAIKVURENTSPKANKEILDEAYVMAGVGSPP 780
DB RKVKVLGGAGFTVYKGIWIPDGENVKIPVAIKVURENTSPKANKEILDEAYVMAGVGSPP 780
QY 781 YVSRLLGICLTSTVOLVTQLMPYGCLLDHRNRCRLGSQDILLNMCQIAKMSYLEVDVR 840
DB YVSRLLGICLTSTVOLVTQLMPYGCLLDHRNRCRLGSQDILLNMCQIAKMSYLEVDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRRPT 900
DB LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRRPT 900
QY 901 HQSDVMSYCVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVTMIMVKWM 960
DB HQSDVMSYCVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVTMIMVKWM 960
QY 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
DB IDSECRPRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020

Db 961 ISECRPRRELVESESRMARDPQRFVVIQNEIDLGPASPLDSTFYSLLEDDMDGLVDA 1020
QY 1021 EBYLVPQOQFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLPESEEAAPRSLAPSEG 1080
Db 1021 EBYLVPQOQFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLPESEEAAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPHDPSPLQRYSEDPVLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPHDPSPLQRYSEDPVLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQPDVPRQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGGAVENPEYLTPO 1200
Db 1141 NQPDVPRQPPSPREGPLPAARPAAGATLERPKTILSPCKNGVVDVFAFGGAVENPEYLTPO 1200
QY 1201 GGAAPOPHPPAFSPAFDNLTYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVPV 1255
Db 1201 GGAAPOPHPPAFSPAFDNLTYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 8
AAY84780
ID AAY84780 standard; Protein; 1255 AA.
XX AC AAY84780;
XX AC
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.
XX Homo sapiens.
XX WO200020579-A1.
XX 13-APR-2000.
XX 01-OCT-1999; 99WO-CA00912.
XX 02-OCT-1998; 98US-0165192.
XX (UYMC-) UNIV MCMASTER.
XX Muller WJ, Siegel PM;
PI WPI: 2000-303768/26.
DR N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.

CC The present sequence represents a SPLICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX Sequence 1255 AA;

Query Match 97.8%; Score 6651; DB 21; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;
QY 1 MELAALCRNGLLLALLPPGAASTOVCTGTDMLKRLPASPEHLRLHLYQGCVVQGNL 60
Db 1 MELAALCRNGLLLALLPPGAASTOVCTGTDMLKRLPASPEHLRLHLYQGCVVQGNL 60
QY 61 ELYLPTNASLSFLQDIQEVQGYVLIAHQVQVPLQRLRIVRGTOIFEDNYALAVLDNG 120
Db 61 ELYLPTNASLSFLQDIQEVQGYVLIAHQVQVPLQRLRIVRGTOIFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLRELOLRSLEILKGGVLIQORNPOLCYQDITLWKDI FHKNNQLA 180
Db 121 DPLNNTTPTVGASPGGLRELOLRSLEILKGGVLIQORNPOLCYQDITLWKDI FHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARGKGLPDTCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARGKGLPDTCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNHSGI CELHCPALVTYNTDTFESMENPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNHSGI CELHCPALVTYNTDTFESMENPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMYIKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAPFESFDGDPASNTAPLOPELOVFTLEETLGYLISAWPDSLP 420
Db 361 IQEFGAGCKKIFGSLAPFESFDGDPASNTAPLOPELOVFTLEETLGYLISAWPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGSGLALIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGSGLALIHNNHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGPTQCVCNSOFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGPTQCVCNSOFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFFCVARC 600
QY 601 PQYIKANSKFITITELPDEEGACQPCPINCETHSCVDLDDKGPAPORASPLTSIVSAVVG 660
Db 601 PSGVKPDLSYMPFWKFPDEEGACQPCPINCETHSCVDLDDKGPAPORASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKVTMRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKVTMRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAVGSP 780
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAVGSP 780
QY 781 YVSRLLGICLSTVOLTPQMPYGLLDHVRNRLGSDLLNMCQIATKMSYLEDVYR 840
Db 781 YVSRLLGICLSTVOLTPQMPYGLLDHVRNRLGSDLLNMCQIATKMSYLEDVYR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWMALESILRRRT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWMALESILRRRT 900
QY 901 HQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPICTIDVTWIMVKCMW 960
Db 901 HQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPICTIDVTWIMVKCMW 960
QY 961 IDSECRPRFRELVSSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYSLLEDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYSLLEDDMDGLVDA 1020
QY 1021 EBYLVPQOQFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLPESEEAAPRSLAPSEG 1080

Db	1021	EEYLVVPOQGFCDPAPAGAGVMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG	1080
Qy	1081	AGSDVFDGDLGMAAGKQLSLPHDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMAAGKQLSLPHDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Qy	1141	NQPDVRPOPSPREGPLPAARPAAGATLERAKTILSPGKNGVVVDVFAFGGAVENPEYLTQP	1200
Db	1141	NQPDVRPOPSPREGPLPAARPAAGATLERPKTILSPGKNGVVVDVFAFGGAVENPEYLTQP	1200
Qy	1201	GGAAPQHPHPPAFSPADNLVYWDQPPPERGAPPPSTFKGTPTAENPEYLGLDVVP	1255
Db	1201	GGAAPQHPHPPAFSPADNLVYWDQPPPERGAPPPSTFKGTPTAENPEYLGLDVVP	1255
RESULT 9			
AAB85458			
ID	AAB85458 standard; Protein; 1255 AA.		
XX	AAB85458;		
AC	XX		
DT	25-SEP-2001 (first entry)		
XX	XX		
DE	Human HER-2/neu protein.		
XX	XX		
KW	Antigen-presenting cell; immunogenic; immune response; HER-2/neu;		
KW	oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
PN	W0200153463-A2.		
XX	XX		
PD	26-JUL-2001.		
XX	XX		
PF	19-JAN-2001; 2001WO-US01850.		
XX	XX		
PR	21-JAN-2000; 2000US-0177545.		
XX	XX		
PA	(CORI-) CORIXA CORP.		
XX	XX		
PI	Cheever MA, Hand-Zimmermann S;		
XX	XX		
DR	WPI; 2001-476112/51.		
DR	N-PSDB; AAH23392.		
XX	XX		
PT	New antigen-presenting cells, useful as vaccines for eliciting or		
PT	enhancing an immune response to HER-2/neu protein, particularly useful		
PT	for treating or preventing cancer, e.g. breast cancer -		
XX	XX		
PS	Claim 2; Page 41-46; 49pp; English.		
XX	XX		
CC	The invention provides an isolated antigen-presenting cell, which		
CC	expresses at least an immunogenic portion of a polypeptide that produces		
CC	an immune response to HER-2/neu protein. The antigen-presenting cells are		
CC	useful as vaccines for eliciting or enhancing an immune response to		
CC	HER-2/neu protein, particularly in treating or preventing malignancies in		
CC	which the HER-2/neu oncogene is associated. Specifically, these are		
CC	useful for treating or preventing cancer, e.g. breast cancer, ovarian,		
CC	colon, lung or prostate cancers. The present sequence represents		
CC	the human HER-2/neu protein (also known as p185 or c-erbB2).		
XX	XX		
SQ	Sequence 1255 AA;		
Query Match 97.8%; Score 6651; DB 22; Length 1255;			
Best Local Similarity 97.6%; Pred. No. 0;			
Matches 1255; Conservative 10; Mismatches 20; Indels 0; Gaps 0;			
Qy	1	MELAAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPEETHDMLRHLHYQGCVVQGNL	60
Db	1	MELAAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPEETHDMLRHLHYQGCVVQGNL	60
Qy	61	ELTYLPTNASLSFLQDIQEVQGVVLIAHNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNG	120

Db	61	ELTYLPTNASLSFLQDIQEVQGVVLIAHNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNG	120
Qy	121	DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDITLWKDIFHKKNOLA	180
Db	121	DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDITLWKDIFHKKNOLA	180
Qy	181	LTLIDTNRGRACHPCSPCKGRCWGESSEDCOSLRTTVCAGGCARCKGPLTDCCHQEC	240
Db	181	LTLIDTNRGRACHPCSPCKGRCWGESSEDCOSLRTTVCAGGCARCKGPLTDCCHQEC	240
Qy	241	AAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Db	241	AAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Qy	301	YNYLSTDVGSCTLVCPHLHQEVTAEDGTORCEKSKPCARVCYGLGHOYIKANSKFITIT	360
Db	301	YNYLSTDVGSCTLVCPHLHQEVTAEDGTORCEKSKPCARVCYGLGMEHUREVRAVTSAN	360
Qy	361	ELEFAGACKI FGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPDSLP	420
Db	361	IQEFAGACKI FGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPDSLP	420
Qy	421	DLVSFQNLQVIRGRILHNGAYSILTQGLGISWGLSLRELGLSGLALIIHNTHLCFVHTV	480
Db	421	DLVSFQNLQVIRGRILHNGAYSILTQGLGISWGLSLRELGLSGLALIIHNTHLCFVHTV	480
Qy	481	PMDQLFRNPHQALLHTANRPEDECVEGEGGLACHQLCARGHCWGPGTQCVCNSOFLRQEC	540
Db	481	PMDQLFRNPHQALLHTANRPEDECVEGEGGLACHQLCARGHCWGPGTQCVCNSOFLRQEC	540
Qy	541	VEECRVLOGLPREYVNNARHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFPCVARC	600
Db	541	VEECRVLOGLPREYVNNARHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFPCVARC	600
Qy	601	POYIKANSKFIGITELPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSIVSAVVG	660
Db	601	PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSII SAVVG	660
Qy	661	ILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQOMILKETEL	720
Db	661	ILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQOMILKETEL	720
Qy	721	RKVKVLGSGAFGTVYKGIWIPDGENVKI PVAIKVLRENTS PKANKETLDEAYVMAGVGP	780
Db	721	RKVKVLGSGAFGTVYKGIWIPDGENVKI PVAIKVLRENTS PKANKETLDEAYVMAGVGP	780
Qy	781	YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSDLLNWCWQIAKGMYSYLEDVR	840
Db	781	YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSDLLNWCWQIAKGMYSYLEDVR	840
Qy	841	LVHRDLAARNVLVKSNNHVKITDPLGLARLLIDETEHADGGKVPKMMALESILRRFT	900
Db	841	LVHRDLAARNVLVKSNNHVKITDPLGLARLLIDETEHADGGKVPKMMALESILRRFT	900
Qy	901	HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPOLLEKGERLPQPPICITIDVYIMVKCM	960
Db	901	HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPOLLEKGERLPQPPICITIDVYIMVKCM	960
Qy	961	IDSECRPRFRELVSFEFSRMARDQRFVJONEDLGPASPLDSTFYRSLLEDDDMGLVDA	1020
Db	961	IDSECRPRFRELVSFEFSRMARDQRFVJONEDLGPASPLDSTFYRSLLEDDDMGLVDA	1020
Qy	1021	EEVLVPQGGFFCDPAPAGGVMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG	1080
Db	1021	EEVLVPQGGFFCDPAPAGGVMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG	1080
Qy	1081	AGSDVFDGLGMAAGKQLSLPHDPSPLQRYSEDPTVPLPSDTGTVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGLGMAAGKQLSLPHDPSPLQRYSEDPTVPLPSDTGTVAPLTCSPQPEYV	1140
Qy	1141	NQPDVAPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVVDVFAFGGAVENPEYLTQ	1200

Db 1141 NOPDVRPQPPSPREGPLPAARAGATLERPKTLSPCKGNVXKDVAFAGGAVENPEYLTRQ 1200
 QY 1201 GGAARQPPPPAFSAFNLVYWDQDPPRGAPPSTFKTPTAENPEYLGLDVVP 1255
 Db 1201 GGAARQPPPPAFSAFNLVYWDQDPPRGAPPSTFKTPTAENPEYLGLDVVP 1255

RESULT 10

AAG88267
 ID AAG88267 standard; Protein; 1255 AA.

AC AAG88267;

DT 11-SEP-2001 (first entry)

XX HER2/neu amino acid sequence.

DE Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
 KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
 KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.

XX Homo sapiens.

OS WO200141787-A1.

PN 14-JUN-2001.

PD 11-DEC-2000; 2000WO-US33591.

PR 10-DEC-1999; 99US-0458299.

XX (EPTM-) EPIMMUNE INC.

XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
 PI Keogh E;

XX WPI; 2001-374995/39.

PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
 PT cellular immune responses for the prevention and treatment of cancer -
 PS Disclosure; Page 15; 199pp; English.

XX The present invention describes isolated prepared HER2/neu epitopes (I).
 CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
 CC culture in vitro and binds to a complex of an epitope (I), bound to a
 CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
 CC and a second epitope and the peptide is less than 50 contiguous amino
 CC acids that have 100% identity with a native peptide sequence of HER2/neu;
 CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
 CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
 CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
 CC and immunostimulant activities, and can be used in vaccines. (I), (II)
 CC and (III) are useful for inducing cellular immune responses for the
 CC prevention and treatment of cancer. (I) and (II) are useful for
 CC monitoring or evaluating an immune response to a tumour-associated
 CC antigen when incubated with a T lymphocyte sample from a patient and
 CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
 CC based vaccines mean that immunosuppressive epitopes that may be present
 CC in whole antigens may be avoided. Selected epitopes may be combined to
 CC enhance immunogenicity. The possible pathological side effects caused by
 CC infectious agents or whole protein antigen is eliminated. The vaccine
 CC provides the ability to direct and focus an immune response to multiple
 CC selected antigens from the same pathogen. Epitope-based anti-tumour
 CC vaccines provides the opportunity to combine epitopes derived from
 CC multiple tumour-associated molecules addressing the problem of tumour-
 CC tumour variability and reducing the likelihood of tumour escape due to
 CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
 CC the exemplification of the present invention.

XX Sequence 1255 AA;

Query Match 97.8%; Score 6651; DB 22; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;
 Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPEHLDMLRHLHYQGQVVOGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPEHLDMLRHLHYQGQVVOGNL 60

QY 61 ELTYLPTNASLSFLQDIQEVQVYLIHNOVRQVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIQEVQVYLIHNOVRQVPLQRLRIVRGTOQLFEDNYALAVLDNG 120

QY 121 DPLNNTTPTVGTASPGGLREQLRLSLTEILKGGVLIQORNPOLCYQDILWKDIFHKNNOLA 180
 Db 121 DPLNNTTPTVGTASPGGLREQLRLSLTEILKGGVLIQORNPOLCYQDILWKDIFHKNNOLA 180

QY 181 LTLIDNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCCAGCARGKGPLTDCCHCEOC 240
 Db 181 LTLIDNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCCAGCARGKGPLTDCCHCEOC 240

QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

QY 301 YNYLSTDVGSCTLVCPHNLHNEVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIGIT 360
 Db 301 YNYLSTDVGSCTLVCPHNLHNEVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIGIT 360

QY 361 EUEFAGCKKIFGSLAFLPESFDGPASNTAPLQPEOLQVFETLEETIGVLIYISAWPDSL 420
 Db 361 IOEFAGCKKIFGSLAFLPESFDGPASNTAPLQPEOLQVFETLEETIGVLIYISAWPDSL 420

QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGSGLALIHNTHLFCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGSGLALIHNTHLFCFVHTV 480

QY 481 PWDQLFRPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSOFLRGQEC 540
 Db 481 PWDQLFRPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSOFLRGQEC 540

QY 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHKDPFCVARG 600
 Db 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHKDPFCVARG 600

QY 601 POYIKANSKFIGITELPDEEGACQCPINCTHSCVDLDDKGPAPAEORASPLTSIVSAVVG 660
 Db 601 PSGVKPDLSYMFIMKFPDEEGACQCPINCTHSCVDLDDKGPAPAEORASPLTSIVSAVVG 660

QY 661 ILLVVVLGVVFGILIKRRQOKIRKVTMRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
 Db 661 ILLVVVLGVVFGILIKRRQOKIRKVTMRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720

QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYMAGVGP 780
 Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYMAGVGP 780

QY 781 YVSRLLGLCTSTVOLVTQLMPYGLLDHVRNRRGLSGQDLNLCWQIAKMSYLEDVDR 840
 Db 781 YVSRLLGLCTSTVOLVTQLMPYGLLDHVRNRRGLSGQDLNLCWQIAKMSYLEDVDR 840

QY 841 LVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGKVPKIMWALESILRRRFT 900
 Db 841 LVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGKVPKIMWALESILRRRFT 900

QY 901 HQSDVMSXGVTWELMTGAKPYDGI PAREIPDLLEKGERLPQPTICTIDVYMWKVM 960
 Db 901 HQSDVMSXGVTWELMTGAKPYDGI PAREIPDLLEKGERLPQPTICTIDVYMWKVM 960

QY 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDEDDMGDLVDA 1020
 Db 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDEDDMGDLVDA 1020

QY 1021 EBYLVPOQGFCTPDPAAGAGMVHRRSSSTRSGGDLTLGLFSEEEAPRSLAPSEG 1080

Db 1021 EYLVFPQQGFFCDPAPAGAGMVMHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYVVDVFAFGGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVYVVDVFAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAPSPAFDNLYYDQDPPRGAPSPSTFKGTPPTAENPEYLGLDVVPV 1255
Db 1201 GGAAPQHPPPAPSPAFDNLYYDQDPPRGAPSPSTFKGTPPTAENPEYLGLDVVPV 1255

RESULT 11

AAE24067
ID AAE24067 standard; Protein; 1255 AA.
XX AAE24067;
DT 23-SEP-2002 (first entry)
DE Human Her-2 protein.
XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
XX Homo sapiens.
OS WO200222636-A1.
PN 21-MAR-2002.
PD 12-SEP-2001; 2001WO-US28572.
PF 15-SEP-2000; 2000US-0663834.
PR (ISIS-) ISIS PHARM INC.
PA Bennett CF, Cowsert LM;
PI WPI; 2002-471192/50.
PX N-PSDB; AAD38904.
XX Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX Example 13; Page 95-107; 116pp; English.
XX The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX Sequence 1255 AA;

Query Match 97.8%; Score 6651; DB 23; Length 1255;
Best Local Similarity 97.8%; Pred. NO. 0;
Matches 1255; Conservative 10; Mismatches 20; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLIALLPFGAASQVCTGTDMLKURLPASPEHLDMLRHLYQGCQVVOGNL 60
Db i MELAALCRWGLLIALLPFGAASQVCTGTDMLKURLPASPEHLDMLRHLYQGCQVVOGNL 60

QY 61 ELTYLPTNASLSFLQDIQEQGYVLIHNOVQVPLQRLIRVRGTQOLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEQGYVLIHNOVQVPLQRLIRVRGTQOLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDILFHKNNOLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDILFHKNNOLA 180
QY 181 LTLIDTNRBRACHPCSPMCKSGRCWGESSEDCOSLRTVTCAGGCARCKGPLTDCCHQEC 240
Db 181 LTLIDTNRBRACHPCSPMCKSGRCWGESSEDCOSLRTVTCAGGCARCKGPLTDCCHQEC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHMQEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHMQEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 ELBFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVEETLEEITGYLYISAWPDSLP 420
Db 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVEETLEEITGYLYISAWPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRELGLALIHNTHLFCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRELGLALIHNTHLFCFVHTV 480
QY 481 PWDQLFRNHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLRQEC 540
Db 481 PWDQLFRNHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLRQEC 540
QY 541 VEECRVLQGLPREYVNHARHCLPCHPECQPNQNGSVTCFGEADOCVACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVNHARHCLPCHPECQPNQNGSVTCFGEADOCVACAHYKDPFPCVARC 600
QY 601 POYIKANSKFIGITELDEEGACQPCINCTHSCVDLDDKGCPEAORASPLTSTVSAVVG 660
Db 601 PSQVKPDLSPYMPKPFDEGACQPCINCTHSCVDLDDKGCPEAORASPLTSTVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRROOKIRKYMRRLLQETELVEPLTPSGAMPNQAOMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRROOKIRKYMRRLLQETELVEPLTPSGAMPNQAOMRILKETEL 720
QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTS PKANKETLDEAYVMAGVSP 780
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTS PKANKETLDEAYVMAGVSP 780
QY 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLGSODLLNWCQIAKGMYSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLGSODLLNWCQIAKGMYSYLEDVR 840
QY 841 LVHRDLAARNLVKSPNNVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRFT 900
Db 841 LVHRDLAARNLVKSPNNVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRFT 900
QY 901 HQSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYMKWM 960
Db 901 HQSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYMKWM 960
QY 961 IDSECRPRELVSEFSRMDPQRFVWIQNEDLGPASPLDSTFYRSLLDEDDMDGLVDA 1020
Db 961 IDSECRPRELVSEFSRMDPQRFVWIQNEDLGPASPLDSTFYRSLLDEDDMDGLVDA 1020
QY 1021 BEYLVPOQGFCDPAPAGAGMVMHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db 1021 BEYLVPOQGFCDPAPAGAGMVMHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYVVDVFAFGGAVENPEYLTPO 1200

Db 1141 NQPDVRQPPSPREGPLPAARPAAGATILRPKTLSPGKNGVVKDVFARGGAVENPEYITPQ 1200
Qy 1201 GGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255
RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX AAE20479;
XX 01-JUL-2002 (first entry)
XX Human Her-2/neu protein.
XX Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
XX Human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Region 1021..1030
XX /note= "Naturally processed HLA-B44-restricted epitope"
XX WO200214503-A2.
XX 21-FEB-2002.
XX 14-AUG-2001; 2001WO-US41733.
XX 14-AUG-2000; 2000US-225152P.
XX 28-SEP-2000; 2000US-236428P.
XX 21-FEB-2001; 2001US-270520P.
XX (CORI-) CORIXA CORP.
XX Hand-zimmermann S, Cheever MA, Foy TW, Lodes MJ, Kalos MD;
XX McNeill PD, Vedvick TS;
XX WPI; 2002-280758/32.
XX N-PSDB; AAD32743.
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
XX prevention and diagnosis of cancer, preferably breast cancer
XX Disclosure; Page 114-117; 129pp; English.
XX The invention relates to an isolated Her-2/Neu polypeptide composition
XX effective for eliciting an immune response. The invention is useful for
XX eliciting an immune response in a patient, where the patient is human
XX leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
XX The composition is useful for the therapy and diagnosis of cancer,
XX preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
XX and other compositions for the diagnosis, prevention and treatment of
XX human malignancies, for stimulating and/or expanding T cells specific for
XX Her-2/Neu polypeptide and for inhibiting the development of cancer in a
XX patient. The invention is useful for stimulating a T cell response in a
XX human patient, as probe or primer for nucleic acid hybridisation to
XX selectively form duplex molecules with complementary stretches of the
XX entire Her-2/Neu gene or gene fragments with complementary stretches of a
XX length gene from a suitable library, and to direct expression of a
XX polypeptide in appropriate host cells. The composition is useful in
XX prophylactic or therapeutic applications and for the treatment of cancer,
XX preferably for the immunotherapy of breast cancer and other Her-2/Neu-
XX associated malignancies. The invention is useful in gene therapy. The
XX present sequence is human Her-2/neu protein.
SQ Sequence 1255 AA;
Query Match 97.8%; Score 6651; DB 23; Length 1255;
1021 EYLVVQQQFFCDDPAPGAGGVMHRRHSSTSGGGDLTLGLEPSEEEAPRPLAPSEG 1080
|

Best Local Similarity 97.6%; Pred. No. 0;
Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPFGAASSTOVCTGDMKRLPASPETHLDMRLHYOGCQVQGNL 60
Db 1 MELAALCRWGLLLALLPFGAASSTOVCTGDMKRLPASPETHLDMRLHYOGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTQLPEDNYALVDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTQLPEDNYALVDNG 120
Qy 121 DPLNNTTPTVGASPGGLRELQLRSITLILKGGVLIQRNPOLCYODTILWKDIFHKQNOLA 180
Db 121 DPLNNTTPTVGASPGGLRELQLRSITLILKGGVLIQRNPOLCYODTILWKDIFHKQNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKCKSGRCWGESSEDCQSLTRTVACAGGACRCKGPLTDCCHQOC 240
Db 181 LTLIDTNRSRACHPCSPMKCKSGRCWGESSEDCQSLTRTVACAGGACRCKGPLTDCCHQOC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHQNQVTAEDGTQCEKCKPCARVCYGLGMHLEHREAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPLNHQNQVTAEDGTQCEKCKPCARVCYGLGMHLEHREAVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFPLESFDGDPASNTAPLOEQLOVPETLEEITGLYVISAWPDSLP 420
Db 361 IQEFAGCKKIFGSLAFPLESFDGDPASNTAPLOEQLOVPETLEEITGLYVISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGIRLHNGAYSLTQLGLISWGLRSLRELGLGSLALIHNTLFCFVHTV 480
Db 421 DLSVFQNLQVIRGIRLHNGAYSLTQLGLISWGLRSLRELGLGSLALIHNTLFCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQICARGHCHGPGPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQICARGHCHGPGPTQCVNCSQFLRGQEC 540
Qy 541 VEBCRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDRPFCVARC 600
Db 541 VEBCRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDRPFCVARC 600
Qy 601 POYIKANSKFITELPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISVAVVG 660
Db 601 PSGVKPDLSYMPIWKPPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISVAVVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTNRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYTNRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSGQDLNWCMIKAGMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSGQDLNWCMIKAGMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITDYYMIWVKWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITDYYMIWVKWM 960
Qy 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDGLPASPLDSTFYRSLLEDGMDLVA 1020
Db 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDGLPASPLDSTFYRSLLEDGMDLVA 1020
Qy 1021 EYLVVQQQFFCDDPAPGAGGVMHRRHSSTSGGGDLTLGLEPSEEEAPRPLAPSEG 1080
|

Db 1021 EYLVPQQGFCPPDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080
 Qy 1081 AGSDVFDGLGMAAKGLQSLPDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGLGMAAKGLQSLPDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
 Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVFAGGAVENPEYLTPQ 1200
 Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVFAGGAVENPEYLTPQ 1200
 Qy 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPBERGAPPSTFKGTPTAENPEYGLDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPBERGAPPSTFKGTPTAENPEYGLDVPV 1255

RESULT 13

AAM51143

ID AAM51143 standard; Protein; 1255 AA.

AC AAM51143;

XX 17-JUN-2002 (first entry)

XX Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
 XX tyrosine kinase; receptor; c-erbB2; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FH Domain 1..653

FT Domain /note= "extracellular domain"

FT Domain 676..1255

FT Domain /note= "intracellular domain"

FT Domain 990..1255

FT Domain /note= "phosphorylation domain"

XX WO200212341-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.

XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
 PT or enhancing an immune response to the protein, has Her-2/neu
 PT extracellular domain fused to Her-2/neu intracellular or
 PT phosphorylation domain

XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
 CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
 CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
 CC in a variety of cancers, including breast, ovarian, colon, lung and
 CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
 CC family of receptor-like glycoproteins. It comprises an extracellular
 CC domain with homology to the epidermal growth factor receptor
 CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
 CC intracellular domain that also shows homology to EGFR. Its
 CC overexpression correlates with a poor prognosis in breast and
 CC ovarian cancers. The invention provides Her-2/neu fusion
 CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
 CC preferred fusion proteins, the extracellular domain of a Her-2/neu
 CC protein is fused to a Her-2/neu intracellular domain or
 CC phosphorylation domain (or its DeltaPp fragment). An immune
 CC response to Her-2/neu protein is elicited or enhanced by
 CC administering the fusion protein in the form of a vaccine, or by
 CC transfecting cells of an animal ex vivo with a nucleic acid
 CC encoding the fusion protein, and delivering the transfected cells
 CC to the animal. The fusion proteins, nucleic acids, and isolated
 CC specific T-cells are useful for inhibiting the development of a
 CC cancer, especially breast, ovarian, colon, lung or prostate cancer
 CC in a patient. T cells that specifically react with a Her-2/neu
 CC fusion protein can be used to remove tumour cells from a sample in
 CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 97.8%; Score 6651; DB 23; Length 1255;
 Best Local Similarity 97.6%; Pred. No. 0;
 Matches 1255; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASPETHLDMRLHYQCQVVQGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASPETHLDMRLHYQCQVVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNOVROVPLQRLRIVRGTLQEDNVALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNOVROVPLQRLRIVRGTLQEDNVALAVLDNG 120

Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLEILKGGVLTQPNQCYQDTILWKDILFHNNQLA 180
 Db 121 DPLNNTTPTVTCASPGGLRELQRLSLEILKGGVLTQPNQCYQDTILWKDILFHNNQLA 180

Qy 181 LTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTVCAAGCARGPLPTDCHEQC 240
 Db 181 LTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTVCAAGCARGPLPTDCHEQC 240

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMQVYKANSKFIGIT 360
 Db 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMHLEHREVRVTSAN 360

Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEITGYLYISAMPDSL 420
 Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEITGYLYISAMPDSL 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRELGSGLALIHNNTHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRELGSGLALIHNNTHLCFVHTV 480

Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQBC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQBC 540

Qy 541 VEECRVLOGLPREVYNAPHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
 Db 541 VEECRVLOGLPREVYNAPHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600

Qy 601 PQYTKANSKFITGELPDEEGACQPCINCTHSCVDLDDKGCAPAEORASPLTISAVVG 660
 Db 601 PSGVKPDLSTYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPAEORASPLTISAVVG 660

Qy 661 ILLVVLGVWFGILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 Db 661 ILLVVLGVWFGILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
 Db 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780

QY 781 YVSRLLGICLTSTVQLVTLMPYGCCLLDHVRENRGRIGSODLLNWCQIAKGSYLEDVR 840
DB 781 YVSRLLGICLTSTVQLVTLMPYGCCLLDHVRENRGRIGSODLLNWCQIAKGSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVKCWM 960
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVKCWM 960
QY 961 IDSECRPRFRELSEFSEMRDPORFVVIQNEIDLGPASPLDSTFYRSLLEDGMDGLVDA 1020
DB 961 IDSECRPRFRELSEFSEMRDPORFVVIQNEIDLGPASPLDSTFYRSLLEDGMDGLVDA 1020
QY 1021 EYLVPOQGFCCPDPAFGAGGVHHRSSSTRSGGDLTLGLEPSEEEAPRSPPLASEG 1080
DB 1021 EYLVPOQGFCCPDPAFGAGGVHHRSSSTRSGGDLTLGLEPSEEEAPRSPPLASEG 1080
QY 1081 AGSDVFDGDLGMAAKGLQSLPTHPDSPLOQRYSEDPTVPLPSETDGVYVAPLTCSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAAKGLQSLPTHPDSPLOQRYSEDPTVPLPSETDGVYVAPLTCSPQPEYV 1140
QY 1141 NQPDVROPQPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVAFGGAVENPEYLTPO 1200
DB 1141 NQPDVROPQPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255
DB 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX AC AAU77114;
XX DT 05-JUN-2002 (first entry)
XX DE Human Her-2/neu polypeptide.
XX KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
XX Hodgkin's lymphoma; T cell therapy.
XX OS Homo sapiens.
XX PN W020213847-A2.
XX PD 21-FEB-2002.
XX PF 13-AUG-2001; 2001WO-US25408.
XX PR 14-AUG-2000; 2000US-0638280.
XX PR 28-SEP-2000; 2000US-0675904.
XX PA (CORI-) CORIXA CORP.
XX PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX WPI; 2002-280741/32.
XX DR N-PSDB; ABK10730.
XX PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
XX polypeptide
PS Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 97.8%; Score 6651; DB 23; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLALLPPGAASQVCTGTDMKRLPASPTHLDMLRHLVGGCVVQGNL 60
DB 1 MELAALCRWGLLALLPPGAASQVCTGTDMKRLPASPTHLDMLRHLVGGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEYGVYLIHNAQVQVPLQRLRIVRGTLFEDNYALVLDNG 120
DB 61 ELTYLPTNASLSFLQDIQEYGVYLIHNAQVQVPLQRLRIVRGTLFEDNYALVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHNQOLA 180
DB 121 DPLNNTPTVTGASPGGLRELQSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHNQOLA 180
QY 181 LTLIDNRSRACHPCSPMKSGSCWGESSEDCSLTRTVCCAGGCARCKGLPTCCHEQC 240
DB 181 LTLIDNRSRACHPCSPMKSGSCWGESSEDCSLTRTVCCAGGCARCKGLPTCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSPCARVCYGLGMOYIKANSKFIGIT 360
DB 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSPCARVCYGLGMOYIKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLQVFEETLEEITGYLIISAWPSLP 420
DB 361 IQBFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLQVFEETLEEITGYLIISAWPSLP 420
QY 421 DLSVFQNLQVIRGRIILHNGAYSILTQGLGISWLGSLRELGLSLALIHNTLHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRIILHNGAYSILTQGLGISWLGSLRELGLSLALIHNTLHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDBCVGEGLACHOLCARGHCWGPPTQCVNCSQFLRQEC 540
DB 481 PWDQLFRNPHQALLHTANRPEDBCVGEGLACHOLCARGHCWGPPTQCVNCSQFLRQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHBECOPONGSVTCFGEADOCVACHYKDRPFCVARC 600
DB 541 VEECRVLQGLPREYVNAHCLPCHBECOPONGSVTCFGEADOCVACHYKDRPFCVARC 600
QY 601 POYIKANSKFIGITELPDEBEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTISAVVG 660
DB 601 PSQVXPDLSPYWKPFDEBEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTISAVVG 660
QY 661 ILLVVLGVVFGILIKRROOKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKETEL 720
DB 661 ILLVVLGVVFGILIKRROOKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKETEL 720
QY 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEITLDEAYMAGVGP 780
DB 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEITLDEAYMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVTLMPYGCCLLDHVRENRGRIGSODLLNWCQIAKGSYLEDVR 840
DB 781 YVSRLLGICLTSTVQLVTLMPYGCCLLDHVRENRGRIGSODLLNWCQIAKGSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900

Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPICKMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCWM 960
Qy 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1020
Db 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1020
Qy 1021 EYLVPOOGFFCDDPAPGAGVMVHRRSSSTRSGGDLTLGLPSEBEEAPRSLAPSEG 1080
Db 1021 EYLVPOOGFFCDDPAPGAGVMVHRRSSSTRSGGDLTLGLPSEBEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLLSPGKNGVVKVDFAFGGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLLSPGKNGVVKVDFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPAPFSPADNLNLYWDDPPERGAPPSPTFGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPHPAPFSPADNLNLYWDDPPERGAPPSPTFGTPTAENPEYGLDVPV 1255

RESULT 15

AAR39568
ID AAR39568 standard; Protein; 1433 AA.
AC AAR39568;
DT 07-FEB-1994 (first entry)
XX Sequence of c-erbB-2 tumour antigen.
XX Tumour antigen; c-erbB-2; glycoprotein.
XX Homo sapiens.
XX W09316185-A.
XX 19-AUG-1993. 93WO-US01055.
XX 05-FEB-1993; 93WO-US01055.
XX 06-FEB-1992; 92US-0831967.
XX (CETU) CETUS ONCOLOGY CORP.
XX (CREA-) CREATIVE BIOMOLECULES INC.
XX Houston LL, Huston JS, Oppermann H, Ring DB;
XX WPI; 1993-272889/34.
XX DR N-PSDB; AAQ46083.
XX New single chain Fv polypeptide binding to C-erbB-2 tumour
XX antigen - for imaging or treating breast or ovarian cancer etc.
XX Disclosure; pages 48-54; 87pp; English.
XX c-erbB-2 refers to a protein antigen expressed on the surface of
XX tumour cells, such as breast and ovarian tumour cells, which is an
XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
XX the location of a stop codon in AAQ46083.
XX Sequence 1433 AA;

Query Match 97.1%; Score 6608; DB 14; Length 1433;
Best Local Similarity 97.1%; Pred. No. 0;

Matches 1218; Conservative 12; Mismatches 25; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLVGGCVVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLVGGCVVQGNL 60
Qy 61 ELYLPTNALSFLQDIOEVQGVYLIAHNVQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELYLPTNALSFLQDIOEVQGVYLIAHNVQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHNQOLA 180
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHNQOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKCKSGRCWGSSEDCOSLRTVCAGGCARCKGPLTCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKCKSGRCWGSSEDCOSLRTVCAGGCARCKGPLTCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSTLVCPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREAVTSAN 360
Db 301 YNYLSTDVGSTLVCPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREAVTSAN 360
Qy 361 ELSFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLBEITCYLYISAMPDLSL 420
Db 361 IQBFAGCKRIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLBEITCYLYISAMPDLSL 420
Qy 421 DLSVFQNLQVIRGILHNGAYSLTLQGLIGISWGLRSLRELGLALIHHTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGILHNGAYSLTLQGLIGISWGLRSLRELGLALIHHTHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQVCNCSQFLRGQEC 540
Qy 541 VEECRVLQGLPREVYNARHCLPCHPECPONGSVTCFPGPADQCVACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREVYNARHCLPCHPECPONGSVTCFPGPADQCVACAHYKDPFCVARC 600
Qy 601 PQYIKANSKFITELPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTISVAVG 660
Db 601 PSGVKPDLSTYMPKIPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTISVAVG 660
Qy 661 ILLVVLGVVFGILIKRROQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRROQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLGSGAPGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVWAGVSP 780
Db 721 RKVKVLGSGAPGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVWAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTLMPYGLLDHVRENRRGLSQDGLNWCMIKAGMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTLMPYGLLDHVRENRRGLSQDGLNWCMIKAGMSYLEDVR 840
Qy 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPICKMALESILRRFT 900
Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPICKMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCWM 960
Qy 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1020
Db 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1020
Qy 1021 EYLVPOOGFFCDDPAPGAGVMVHRRSSSTRSGGDLTLGLPSEBEEAPRSLAPSEG 1080
Db 1021 EYLVPOOGFFCDDPAPGAGVMVHRRSSSTRSGGDLTLGLPSEBEEAPRSLAPSEG 1080

Qy	1081	AGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NOPDVRPOPSPREGPLPAARAGATLERAKTILSPGKNGVWKDVFAFGGAVENPEYLTPO	1200
Db	1141	NOPDVRPOPSPREGPLPAARAGATLERAKTILSPGKNGVWKDVFAFGGAVENPEYLTPO	1200
Qy	1201	GGAAPQPHPPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255
Db	1201	GGAAPQPHPPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 08:41:45
Job time : 42.9774 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds

(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-579-593-12

Perfect score: 6803

Sequence: 1 MELAAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVFPV 1255

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mbc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6147	90.4	1259	6	O18735 canis famil
2	3112	45.7	1209	11	Q9X70 rattus norv
3	3087	45.4	1210	11	Q9EP98 mus musculu
4	2698	39.7	1165	13	Q9YH40 Q9YH40 xiphophorus
5	2680.5	39.4	1137	13	Q9W6F6 Q9W6F6 gallus gall
6	2280	33.5	1328	13	P79754 P79754 fugu rubrip
7	2028.5	29.8	1433	5	Q9BHH9 Q9BHH9 anopheles g
8	1871	27.5	419	4	Q9UK79 Q9UK79 homo sapien
9	1739	25.6	367	11	Q8R2X1 Q8R2X1 mus musculu
10	1697.5	25.0	412	4	Q8WYV0 Q8WYV0 homo sapien
11	1690	24.8	729	15	Q86712 Q86712 avian rous-
12	1688	24.8	567	15	Q86714 Q86714 avian rous-
13	1623.5	23.9	962	15	Q64895 Q64895 avian eryth
14	1615	23.7	545	15	Q85468 Q85468 avian eryth
15	1478.5	21.7	655	11	Q9WVF5 Q9WVF5 mus musculu
16	1462.5	21.5	643	11	Q9ERV6 Q9ERV6 mus musculu

17	1273	18.7	1193	5	Q9Y1X8	Q9Y1X8 ephydatia f
18	1197.5	17.6	1368	5	Q23821	Q23821 caenorhabdi
19	1176	17.3	1717	5	Q26566	Q26566 schistosoma
20	1126	16.6	527	13	Q90836	Q90836 gallus gall
21	1001.5	14.7	478	11	Q9ESE0	Q9ESE0 rattus norv
22	912.5	13.4	599	13	Q9FSH2	Q9FSH2 gallus gall
23	906	13.3	165	4	Q14256	Q14256 homo sapien
24	812	11.9	176	11	Q923V5	Q923V5 rattus norv
25	806.5	11.9	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	Q8SZW1	Q8SZW1 drosophila
27	754.5	11.1	311	13	Q9Y162	Q9Y162 xiphophorus
28	734	10.8	331	4	Q9BUD7	Q9BUD7 homo sapien
29	727.5	10.7	1362	13	Q9FVZ4	Q9FVZ4 xenopus lae
30	725	10.7	1671	5	Q9NJV5	Q9NJV5 biophalari
31	723	10.6	149	6	Q9BG66	Q9BG66 oryctolagus
32	692	10.2	1418	13	Q93457	Q93457 scophthalmu
33	689.5	10.1	1368	13	Q8UW85	Q8UW85 paralicthty
34	678.5	10.0	1369	13	Q8UW86	Q8UW86 paralicthty
35	668.5	9.8	1472	5	Q9U5A8	Q9U5A8 bombyx mori
36	658	9.7	1358	13	Q73798	Q73798 xenopus lae
37	648	9.5	1412	13	Q8UW84	Q8UW84 paralicthty
38	647.5	9.5	1418	13	Q8UW83	Q8UW83 paralicthty
39	634	9.3	1245	13	Q9YGH8	Q9YGH8 scophthalmu
40	626.5	9.2	2144	5	Q9VD94	Q9VD94 drosophila
41	622	9.1	1371	11	Q9QVM4	Q9QVM4 rattus sp.
42	595	8.7	987	11	Q91YMO	Q91YMO mus musculu
43	590	8.7	935	4	Q96L35	Q96L35 homo sapien
44	590	8.7	987	11	Q99ME2	Q99ME2 mus musculu
45	587.5	8.6	1036	4	Q07912	Q07912 homo sapien

ALIGNMENTS

RESULT 1

O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE ErB-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbB-2 from canine mammary gland";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1;
DR HSSP; P11362; IFCK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4CD46 CRC64;

Query Match		90.4%;	Score 6147;	DB 6;	Length 1259;
Best Local Similarity		90.0%;	Pred. No. 0;		
Matches 1134;		Conservative 46;	Mismatches 74;	Indels 6;	Gaps 2;
Qy	1	MELAAACRGLLLALLP	PGAASTQVCTGDMKRLRLPASBETHLDMRLHLYQGCVVQGNL	60	
Db	1	MELAAACRGLLLALLP	PSGAAGTQVCTGDMKRLRLPASBETHLDMRLHLYQGCVVQGNL	60	
Qy	61	ELTYLPTNASLFLQDI	QEOGVVLIATHNQVROLRIVRGTOLFEDNYALAVLDNG	120	
Db	61	ELTYLPANASLFLQDI	QEOGVVLIATHSQVROIPLQRIVRGTOLFEDNYALAVLDNG	120	
Qy	121	DPLNNTTPTVGTAS	PGGLRELQRLSLTEILKGGVLIQRNPOLCYQDITILWKDIFPHKNQLA	180	
Db	121	DPLEGGIPAGAAQ	AGGLRELQRLSLTEILKGGVLIQRNPOLCHQDITILWKDIFPHKNQLA	180	
Qy	181	LTLIDNRSRACHPC	SPMKGRSCWGESSEDCOSLRTTVCGAGCARCKGPLPTDCCHQC	240	
Db	181	LTLIDNRSFACPC	SPACKDAHCWGAASSGDCOSLRTTVCGAGCARCKGPPPTDCCHQC	240	
Qy	241	AAGCTGPKHSDCL	ACLHFNHSGICELHCPALVTYNTDTFESMNPBGRYTFFGASCVTAC	300	
Db	241	AAGCTGPKHSDCL	ACLHFNHSGICELHCPALVTYNTDTFESMNPBGRYTFFGASCVTSC	300	
Qy	301	YNYLSTDVGSCT	LVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT	360	
Db	301	YNYLSTDVGSCT	LVPLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN	360	
Qy	361	ELFACGKIKFGL	ATLAFPSFDGPASNTAPLOEQLVPETLEETIGYLIYSAMDPSLP	420	
Db	361	IOEFACGKIKFGL	ATLAFPSFDGPASNTAPLOEQLRVFEALEEITGYLIYSAMDPSLP	420	
Qy	421	DLISVFONLQVIR	GRILHNGAYSILTLQGLGTSWGLSLRELGLSLALIHNNHLCFVHTV	480	
Db	421	NLSVFONLQVIR	GRVLDHNGAYSILTLQGLGTSWGLSLRELGLSLALIHNNHLCFVHTV	480	
Qy	481	PMDQLFRNPHQAL	LHTANRPEDCEVCGEGACHOLCARGHCGPPTQCVNCSOFLRGQEC	540	
Db	481	PMDQLFRNPHQAL	LHSANRPEECVCGEGACYP-CAHGHCGPPTQCVNCSOFLRGQEC	539	
Qy	541	VEECRYVQLGL	PRYVNAHCLPCHPECPONGSVTCFPGADOCVACAHYKDPFPCVARC	600	
Db	540	VEECRYVQLGL	PREYVNDYCLPCHSECQPNQSVTCFPGSEADQCVACAHYKDPFPCVARC	599	
Qy	601	POYIKANSRFIG	ITELPDEGACQPCPINCTHSCVDLDDKGCAPAEORASPLTISVAVVG	660	
Db	600	PSGVKDLSPMT	WKFADEGTQCPINCTHSCADLDEKGCAPAEORASPVTSIIAAVVG	659	
Qy	661	ILLVVLGVVFG	ILIKRRQKIRKYMRLLOETELVEPLTPSGAMPNQAMRILKETEL	720	
Db	660	ILLVAVVGL	ILIKRRQKIRKYMRLLOETELVEPLTPSGAMPNQAMRILKETEL	719	
Qy	721	RKVKVLGSGAG	FTYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP	780	
Db	720	RKVKVLGSGAG	FTYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP	779	
Qy	781	YVSRLLGICLT	SVQLVTQMLPMYGCILLDHVRENRLGSLDNLNWCQIAKMSYLEDVR	840	
Db	780	YVSRLLGICLT	SVQLVTQMLPMYGCILLDHVREHRLGSLDNLNWCQIAKMSYLEDVR	839	
Qy	841	LVRDLAARNVL	VKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMMALESILRRRFT	900	
Db	840	LVRDLAARNVL	VKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMMALESIPRRFT	899	
Qy	901	HQSDVMSYGV	TWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCM	960	
Db	900	HQSDVMSYGV	TWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCM	959	
Qy	961	IDSECRPRFEL	VSFEFSRMARDPQRFVIONEDLGASPLDSTFYRSLLEDGMDGLVDA	1020	
Db	960	IDSECRPRFEL	VAEFSRMARDPQRFVIONEDLGASPLDSTFYRSLLEDGMDGLVDA	1019	
Qy	1021	EYLVVPOQGF	FCPPAPAGAGMWHRRHSSTSGGGDLTLGLLEPSEEEAPRSLAPSEG	1080	

Db	1020	EVVLVPOQGF	FCPEPTGAGGTAHRRHSSTNRNGGELTLGLEPSEEEPPKSLAPSEG	1079	
Qy	1081	AGSDVFDGDL	GMAAGLQSLTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV	1140	
Db	1080	AGSDVFDGDL	GMAAGLQSLTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV	1139	
Qy	1141	NQPDVTPQPPS	PREGPLPAARPAAGATLER-----AKTSPGKNGVVDVFAFGCAVENPE	1195	
Db	1140	NOPEVWPQPL	ALEGPLPSPRPAAGATLERPKTLSPGKNGVVDVFAFGCAVENPE	1199	
Qy	1196	YITPQGAAPQ	PHPPAFSPAFDNLNLYWDDPPERGAPSTFTKGTPTAENPEYLGDPV	1255	
Db	1200	YLAPGRAAPQ	PHPPAFSPAFDNLNLYWDDPPSERGSPSTFEGTPTAENPEYLGDPV	1259	
RESULT 2					
Q9QX70	ID	Q9QX70	PRELIMINARY;	PRT;	1209 AA.
AC	Q9QX70;				
DT	01-MAY-2000	(TrEMBLrel. 13, Created)			
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)			
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)			
DE	Epidermal growth factor receptor.				
GN	EGFR.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
RN	[1] SEQUENCE FROM N.A.				
RC	STRAIN=FISHER; TISSUE=LIVER;				
RX	MEDLINE=90258888; PubMed=2342466;				
RA	Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,				
RA	Karp H.S.;				
RT	"A truncated, secreted form of the epidermal growth factor receptor is				
RL	encoded by an alternatively spliced transcript in normal rat tissue.";				
RL	Mol. Cell. Biol. 10:2973-2982(1990).				
RN	[2] SEQUENCE FROM N.A.				
RC	STRAIN=FISHER; TISSUE=LIVER;				
RA	Petch L.A.;				
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.				
RN	[3] SEQUENCE FROM N.A.				
RC	STRAIN=FISHER; TISSUE=LIVER;				
RA	Guttridge K., Dawson T.L., Karp H.S.;				
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; M37394; AAF14008.1;				
DR	HSSP; P11362; 1FGK.				
DR	InterPro; IPR000494; EGFR_L_domain.				
DR	InterPro; IPR000719; Euk_pkinase.				
DR	InterPro; IPR002174; Furin-like.				
DR	InterPro; IPR001245; Tyr_pkinase.				
DR	Pfam; PF00757; Furin-like; 1.				
DR	Pfam; PF00069; pkinase; 1.				
DR	Pfam; PF01030; Recep_L_domain; 2.				
DR	PRINTS; PR00109; TYRKINASE.				
DR	ProDom; PD000001; Euk_pkinase; 1.				
DR	SMART; SM00261; FU; 3.				
DR	SMART; SM00219; TyrKc; 1.				
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.				
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.				
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.				
KW	ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.				
SQ	SEQUENCE 1209 AA; 134891 MW; 96FE7F6CC1B7773 CRC64;				
Query Match 45.7%; Score 3112; DB 11; Length 1209;					
Best Local Similarity 50.0%; Pred. No. 3e-226;					
Matches 639; Conservative 164; Mismatches 359; Indels 116; Gaps 28;					
Qy	3	LAALCRWGLLLALLP	PGA-ASTQVCTGDMKRLRLPASBETHLDMRLHLYQGCVVQGNLE	61	

Db 15 LAALCAAG-----GALBEKKVCOGTSMRLTQLTGTFDEHFLSLQRMFNNECVVLGNLE 66
Qy 62 LTYLPTNASLFLQDQEOGVYVLIHANVQVPLQBLRIVRGTLQEDYVALAVLDNGD 121
Db 67 ITVVQRNDYLSFKTQEVAGYVLIATNVERPLENLQIRGNALYENTYALAVLSN-- 124
Qy 122 PLNNTPPVTCASPGGLRELQRLSLTEILKGVLQIRNPOLCYODTILWKLDFHKNQAL 181
Db 125 -----YGTNKTGLRELPMRLQELIIGAVFSNPIICNMETIQMDIV-QDVFLSN 175
Qy 182 TLDITNRS-RACHPCSMKSGSCWGESSEDCQSLTRTVACGCA-RCKGPLPTDCCHEQ 239
Db 176 MSMDVQRHLTGCPKDCPCNGSCWGRBENCQKLTIIICAQCSCRRGRSPSDCCNQ 235
Qy 240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 299
Db 236 CAAGCTGPRSDCLVCHRFDREATCKDTCPLMLYNPTTYQMDVNPEGKYSFGATCVKCK 295
Qy 300 PYNLYSTDVSGSTLVCPLHMQEVTAEQTCQKCKKPCARVCVGLCMQVYKANSKFIGI 359
Db 296 PRNVVYTDHSGSVRACGPDYEV-EEDGVSKKCKDGPCKVNGVIGIEFK-DTUSINA 353
Qy 360 TELE-FAGCKKIFGSLAFIPESFDGPASNTAPLOPEQLQVFTLBEITGLYISAMPDS 418
Db 354 TNIKFKYCTAISGDLHLPLVAFKGSFTRTPPLDPRELEILTKVTEITGFLIQAOPEN 413
Qy 419 LPDLSVFQNLQVIRGLHNGAYSLTLOGLISWGLRSRLGSLGALIHHTHLCFVH 478
Db 414 WTDLHAFENLEIIRGTRTKHQHQSLSLAVGLNITSLGRLSLEKISDGVITISGRNLICAN 473
Qy 479 TVPWDLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCPGPGTCVNCVSQFLRGQ 538
Db 474 TINWKLFTPNQTKIMNRAEKDKATNHVCNPLCSGCGGPGPTDCVSCQNVSRGR 533
Qy 539 ECVEECRVLOGLPREVYNARHCLUPCHPEQOPQNGSVTCFGEADQCACAHYKDPFCVA 598
Db 534 ECVDKCNILEGEPREFVENSECQCHPECLPOTMNTICTGRGPDNCKCAHYVDGPHCVK 593
Qy 599 RCPQYIKANSKFIGITLPEDEGACOPCPCINCHTSCVDLDDKCPAEQASRP-LTISVA 657
Db 594 TCPSGIMGENTLV-WKFDANNVCHLCHANCTYGCAGPGLKGC--QOPEGPKIPSIATG 650
Qy 658 VVGILLVYVGVVFGI-LIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAOWRLK 716
Db 651 IVGGLLFIVV-VALGIGLFNRRLQVRLTLRLQLRELVEPLTPSGAPNQAHLRIK 709
Qy 717 ETELKRVKVLGSAFTGYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYMAG 776
Db 710 ETEFKIKVLGSAFTGYKGLWIPGEKVIPVAIKELREATSPKANKEILDEAYMAS 769
Qy 777 VGSPYVRLIGLCLTSTVOLVTQLMYPYGLLDHVRENRLGSLQDLLNMCQIAKMSYL 836
Db 770 VDNPHVRLGLCLTSTVOLVTQLMYPYGLLDYVREHKONIGSQYLLNMCVQIAKGMNYL 829
Qy 837 EDVRLVHRLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILR 896
Db 830 EDRRLVHRLAARNLVKTPQHVKITDFGLAKLLGAEKEYRAEGKVPKMALESILH 889
Qy 897 RFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMV 956
Db 890 RIYHQSDVMSYGVTVWELMTGSKPYDGIPIASEISSILEKGERLPQPPICITIDVYIMV 949
Qy 957 KCMWIDSECRPRFELVSEFSRMARDPQFVVIQ-NEDLGPASPLDSTYRSLLLEDDMG 1015
Db 950 KCMWIDASRPKRELILEFSKWARDPQRYLYIQGDERMHLPSPTDSNFYALMEEDME 1009
Qy 1016 DLVDAEYLVPOQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLPSEEPASPL 1075
Db 1010 DVVDAEYLVPOQGF-----NSPST-----SRPL 1035
Qy 1076 APSEGAGSDVFDGLGMGAAGLQSLPHTDPSPLQRYSDPTVPLPSET--DGYVAPLTC 1133
Db 1036 LSSLSANSN-----SSTVACINRNGSCRVKDEAFQRYSSDPTSVLTEDNIDDTFL----- 1086

Qy 1134 SPOEYVYVQPDVVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVVKVFAFGGAVEN 1193
Db 1087 -PVPEYVINO-SVPRKPAQSVQNPVYHNOPLHP-----APGRDLHYQN--PHSNVSN 1134
Qy 1194 PEYL-TPOGAAQPPHPPPAFSPAFONLYYWDQ-----DP-----PERCAPPST 1236
Db 1135 PEYLNTAQ-----PTCLSSGPFSSALWIKQSHQMSLDNPDYQODFFPKAKPNGI 1185
Qy 1237 FKGTPTAENPEYLGLDVP 1254
Db 1186 FKG-PTAENAEYLRVAPP 1202
RESULT 3
Q9EP98
ID Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RA "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RA "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAG28045.1; -
DR EMBL; AF275366; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG24386.1; -
DR HSSP; P11362; 1FGK.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000345; CytC heme bind.
DR InterPro; IPR000494; EGFR L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TyRK; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

QY 549 GLPREY-VNARHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVACP----- 601
Db 507 SLPLRLYSVDSKTCGDCHQCKD-----FCYGNEDNCGSCMVNKGDFCVAECPTTKHAM 561
QY 602 --QYKANSKFTGI-----TELPD----- 618
Db 562 NGTCINCHKTCVGCRCRPTDIAPDGCISCDKALIGSDAKIERCLMKDESCPDGYSDYVL 621
QY 619 -EEG----- 621
Db 622 QEEGPLQLSGRAVCRCKCHPRCKKCTGYGFHEQFCOECTGYKKEQCEDECPDFYANEE 681
QY 622 --ACQPCPINCT-----HSCVDL-----DD-----KCPAEQ----- 646
Db 682 TRICLPCHQECRGCHGLGDHHECRNLKLFEGDPYDNATFTVCNCPASHPKYRFPQEA 741
QY 647 -----RASPLTSIVSAVVGILVVLVGVVFGI---LKRROQKIRKTYM 687
Db 742 GKIGPYCSADSMQSLRIEPTQVKIVMGSMALILLCVVFGIAFVLSRHKNNKDAVKM 801
QY 688 RRLLOETELVEPLTPSGAMPNQOMRILKETELURKVKVLGSGAGFTYKGIWIPDGENVK 747
Db 802 TMALAGCEDSEPLRPSNVGNLTKLRIKEAETIRRGVGLMGAFGRVFKGVMPGESVK 861
QY 748 IPVAIKVLRENTSPKANKETLDEAYVMAGVSPVSRLLGICLTSTVQLVTOLMPYGCLL 807
Db 862 IPVAIKVLEMSGESSEKFELEAYIMASVEHPNLKLLKLAACVMTSQMLITQMLPGCLL 921
QY 808 DHVRENRGLSGDQLNWCQIAKGMSTYLEDVRLVHRDLAARNVLKPSNVHVKITDFGLA 867
Db 922 DYVRNNKDKIGSKALLNWSQIARGWAYLEERLVRDLAARNVLQTPSCVKITVFGLA 981
QY 868 RLIDIDETEYHAGGKVPKIMMALESILRRRPTHQSDVMSYGVTVVWELMTFGAKPYDGIP 927
Db 982 KLDFDSDEYRAAGGKMPIKWLAECTIRHRVFTSKSDWAFGITIWELLTYGARPYENVP 1041
QY 928 AREIPDLKEGERLPOPPICITIDYVIMVKWCMIDSECRPRFELVSEFRMARDPORFV 987
Db 1042 AKDVPELTETGHKLPPQDICSLDVYCILSCWLDADARTFKQLAETFAEKARDPGRYL 1101
QY 988 VIONEDLGPAASPLDSTFYRSLLEDDMDGLV----- 1018
Db 1102 MI-----PGDKFMRPLSYTNQDEKDLIRTLAPVMAAAAAAAGASNVDPSTIA 1152
QY 1019 DAEYLVPOQGFPCPDAPCAGGMVHRHRSSTSRSGGDLTLGLPSEEEAPRS----- 1073
Db 1153 ETDEYLOPKTRPSIMLPGPSA-----VEPS-DEMPKSLRYCK 1188
QY 1074 -PLAP---SEGAGSDVFDGDLGMAAGLQSLPHDPSPLQRYSEDPTVPLPSETDGYVA 1129
Db 1189 DPLKPDDETGHGKEV-----GVGGIR-----LNLPLDEDDYLM 1222
QY 1130 PLTCSQPOEYVNPQDVRPQPPSPREGPLPAARPAATLERAKTLPSPKNGVKDVFAGG 1189
Db 1223 P-TCQSQ---NQS-----TPG-----YMDLIGVPA 1243
QY 1190 AVENPEYL-----TPQGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPSTFKGT 1240
Db 1244 SVDNPEYLMGSTQAIAGLAGSGMG---PHTPP-----PPNTPNGM 1280
QY 1241 PTAENPE 1247
Db 1281 PTHQHSQ 1287

RESULT 8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Herstatie.

GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-Like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;
Query Match 27.5%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 4.5e-133;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMLKRLPASPEHLMLRHLVQGCQVVGNNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMLKRLPASPEHLMLRHLVQGCQVVGNNL 60
QY 61 ELTYLPTNASLSFQDIQEVQGVLIHQNQVROVPLORLIRVGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFQDIQEVQGVLIHQNQVROVPLORLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
QY 181 LTLIDNTRGRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAAGCARGKPLPTDCCHQEC 240
Db 181 LTLIDNTRGRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAAGCARGKPLPTDCCHQEC 240
QY 241 AAGCTGPKHSDDLCLAHFNHSGICELHCPALVYNTDTFESMPNPEGRTYTFGASCVTACP 300
Db 241 AAGCTGPKHSDDLCLAHFNHSGICELHCPALVYNTDTFESMPNPEGRTYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHMQEVTAEQGTORCEKSKPCARVCYGL 345
Db 301 YNYLSTDVGSCTLVCPHMQEVTAEQGTORCEKSKPCARVTHSL 345

RESULT 9
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027080; AAH27080.1; -
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; OBE03395F9E101B0 CRC64;

Query Match 25.6%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 3.6e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 889 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICT 948
DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICT 60

QY 949 IDVYIMVWKWMDIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1008
DB 61 IDVYIMVWKWMDIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 120

QY 1009 LEDDDMGDLVDABEYLVPQGGFFCPDPAPGAGMWHRRSSSTRSGGDLTLGLPSEE 1068
DB 121 LEDDDMGDLVDABEYLVPQGGFFCPDPAPGAGMWHRRSSSTRSGGDLTLGLPSEE 180

QY 1069 EAPRSLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPVLPSETDGYV 1128
DB 181 EAPRSLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPVLPSETDGYV 240

QY 1129 APLTCSQPQEVYNQPDVVRPQPPSPREGPLPAAPAGATLERAKTLPKNGVVKDVFAFG 1188
DB 241 APLTCSQPQEVYNQPDVVRPQPPSPREGPLPAAPAGATLERAKTLPKNGVVKDVFAFG 300

QY 1189 GAVENPEYLTPOGGAAPQHPHAFSPAFDNLVWDODPPERCAPSTFKGTPTAENPEY 1248
DB 301 GAVENPEYLTPOGGAAPQHPHAFSPAFDNLVWDODPPERCAPSTFKGTPTAENPEY 360

QY 1249 LGLDVVP 1255
DB 361 LGLDVVP 367

RESULT 10

Q8WYV0
ID Q8WYV0 PRELIMINARY; PRT; 412 AA.
AC Q8WYV0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN PP3659.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF318349; AAL55856.1; -
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 25.0%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 5.8e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 889 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICT 948
DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICT 60

Db 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICT 60

QY 949 IDVYIMVWKWMDIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1008
DB 61 IDVYIMVWKWMDIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 120

QY 1009 LEDDDMGDLVDABEYLVPQGGFFCPDPAPGAGMWHRRSSSTRSGGDLTLGLPSEE 1068
DB 121 LEDDDMGDLVDABEYLVPQGGFFCPDPAPGAGMWHRRSSSTRSGGDLTLGLPSEE 180

QY 1069 EAPRSLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPVLPSETDGYV 1128
DB 181 EAPRSLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPVLPSETDGYV 240

QY 1129 APLTCSQPQEVYNQPDVVRPQPPSPREGPLPAAPAGATLERAKTLPKNGVVKDVFAFG 1188
DB 241 APLTCSQPQEVYNQPDVVRPQPPSPREGPLPAAPAGATLERAKTLPKNGVVKDVFAFG 300

QY 1189 GAVENPEYLTPOGGAAPQHPHAFSPAFDNLVWDODPPERCAPSTFKGTPTAENPEY 1248
DB 301 GAVENPEYLTPOGGAAPQHPHAFSPAFDNLVWDODPPERCAPSTFKGTPTAENPEY 360

QY 1221 YWYD-QDPPER-----GAPPSTFKGTPTAEN 1245
DB 361 WYWTQCEPEGVRRSPDVSSSGSREGLTSAKIKRWEGPPTTSRGTCARN 410

RESULT 11

O86712
ID O86712 PRELIMINARY; PRT; 729 AA.
AC O86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
ON NCBI_TaxID=11950;
RX [1]
RP SEQUENCE FROM N.A.
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnsson A., Beug H.;
RT "Retroviral capture of c-erbA proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -
DR HSSP; P03322; 1A6S.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR004028; Retro_M.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro_M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914BEFD63 CRC64;

Query Match 24.8%; Score 1690; DB 15; Length 729;
Best Local Similarity 54.2%; Pred. No. 5.1e-119;
Matches 354; Conservative 75; Mismatches 122; Indels 102; Gaps 15;

QY 569 PQNGSVTCFGEADOCVCAHYKDPFCVARGCPQVTKANSKFGITELDEGACQCPPI 628
DB 141 PRETATPKTGP--DHCMKCNHFDHPCVKACPAVLGENDTL-VWKYADANAVCQCHP 197

QY 629 NCHTSCVDLDDKCPAQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKTYM 687
DB 1 NCHTSCVDLDDKCPAQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKTYM 687

```
Db 198 NCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVWVGLGIGLYLRRR-HIVRKRTL 253
Qy 688 RLLLOETELVEPLTPSGAMPNQAMRILKETELRKVKVLGSGAGFTYVKGWIWPDGENVK 747
Db 254 RLLLOETELVEPLTPSGAMPNQAMRILKETELRKVKVLGSGAGFTYVKGWIWPDGENVK 313
Qy 748 IPVAIKVLRNTPSKANKELIDELAYVMAGVSPVSRLLGICLTSTVQLVTLQMPYGCGL 807
Db 314 IPVAIKELREATSPKANKELIDELAYVMASVDNPRVCRLLGICLTSTVQLVTLQMPYGCGL 373
Qy 808 DHVRENRLGRLGQDLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLA 867
Db 374 DYIREHNDKNGSQVLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLA 433
Qy 868 RLIDIDETEHADGCKVPIKMALESILRRFTHQSDVMSVGVTVWELMTFGAKPYDGIP 927
Db 434 KLGADEKEHAEAGCKVPIKMALESILHRIYTHQSDVMSVGVTVWELMTFGSKPYDGIP 493
Qy 928 AREIPDLLEKGERLPQPPICITIDVYIMVWKMWIDSECRPRFRELSEFSRMARDPQRFV 987
Db 494 ASEISSVLEKGERLPQPPICITIDVYIMVWKMWIDSECRPRFRELSEFSRMARDPQRFV 553
Qy 988 VIQ-NEDLGASPLDSTFYRSLLEDDMDGLVDAAEYLVPOQGFPCDPAPGAGGMVHR 1046
Db 554 VIOGDERMHLPSPTDSKPYRTLMEEDMEDIVDAEYLVPHQGF-----598
Qy 1047 HRSSTSTRSGGDLTLGLEPSEEAAPRSP-----APSEGAGSDVFDGLGMGAAGKGLQSL 1101
Db 599 -NSPST-----SRTPLLSLSATSNNSATNCID-----RNGOGH 631
Qy 1102 PTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPOEYVNPQDVPRPQPSREGPLPA 1159
Db 632 PVREDSFVQRYSDPTGNFLEESIDDGFL-----PAPEYVNO--LMPKKPS-----675
Qy 1160 ARPAGATLERAKTUSPQNGVVKDVF-----APFGAVENPEYL 1197
Db 676 -----TAMVQNOIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 715

RESULT 12
Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714; MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynocheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSP; P11362; IFCG.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TRKINASE.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON_TER 1
```

```
SEQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;
Query Match 24.8%; Score 1688; DB 15; Length 567;
Best Local Similarity 54.8%; Pred. No. 4.9e-119;
Matches 353; Conservative 73; Mismatches 116; Indels 102; Gaps 15;

Qy 578 GREADQCVACAHYKPPFCVVARCPQYIKANSKFGITELPDEEGACQPCPINCTHSCVDL 637
Db 1 GP--DHCMKAHFIDGPHCVKACPAAGVLGENDTL-VWKYADANAVCQLCHPNCCTRGCKGP 57
Qy 638 DDGCPABQASPLTISVSAV--GILLVVVLGVVFGILIKRQOKIKRYTMRRLLOETEL 696
Db 58 GLEGCP---NGSKTPSIAAGVVGGLLCLVWVGLGIGLYLRRR-HIVRKRTLRLQLREL 113
Qy 697 VEPLTSGAMPNQAMRILKETELRKVKVLGSGAGFTYVKGWIWPDGENVKIPVAIKVLR 756
Db 114 VEPLTSGEAPNQAMRILKETELRKVKVLGSGAGFTYVKGWIWPDGENVKIPVAIKVLR 173
Qy 757 ENTSPKANKELIDELAYVMAGVSPVSRLLGICLTSTVQLVTLQMPYGCGLDHYRENRR 816
Db 174 EATSPKANKELIDELAYVMASVDNPRVCRLLGICLTSTVQLVTLQMPYGCGLDYIREHKN 233
Qy 817 LGSODLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLDIDETE 876
Db 234 IGSQVLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLAKLGADEKE 293
Qy 877 YHAGGKVPKIKMALESILRRFTHQSDVMSVGVTVWELMTFGAKPYDGIPAREIPDLLE 936
Db 294 YHAGGKVPKIKMALESILHRIYTHQSDVMSVGVTVWELMTFGSKPYDGIPASEISSVLE 353
Qy 937 KGERLPQPPICITIDVYIMVWKMWIDSECRPRFRELSEFSRMARDPQRFVIO--NEDLG 995
Db 354 KGERLPQPPICITIDVYIMVWKMWIDSECRPRFRELSEFSRMARDPQRFVIOGDERMH 413
Qy 996 PASPLDSTFYRSLLEDDMDGLVDAAEYLVPOQGFPCDPAPGAGGMVHRHRSSTSTRSG 1055
Db 414 LPSPTDSKPYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST---454
2y 1056 GGDLTGLEPSEEAAPRSP-----APSEGAGSDVFDGLGMGAAGKGLQSLPTHDPSPLO 1110
Db 455 -----SRTPLLSLSATSNNSATNCID-----RNGOGHPVREDSFVQ 491
Qy 1111 RYSEDPTVPLPSET--DGYVAPLTCSPOEYVNPQDVPRPQPSREGPLPAARPAGATLE 1168
Db 492 RYSDPTGNFLEESIDDGFL-----PAPEYVNO--LMPKKPS-----526
Qy 1169 RAKTUSPQNGVVKDVF-----APFGAVENPEYL 1197
Db 527 -----TAMVQNOIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 566

RESULT 13
Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895; MEDLINE=90206603; PubMed=1969616;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
GN GAG, V-ERB-A, V-ERB-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Bruskina A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B.";
RL Oncogene 5:15-24(1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -.
```

```

DR EMBL; X52211; CAA36459.1; JOINED.
DR HSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stdhmrn_receptor.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000001; Euk_pkinase; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; Tyrc; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match 23.9%; Score 1623.5; DB 15; Length 962;
Best Local Similarity 51.2%; Pred. No. 8.2e-114;
Matches 354; Conservative 73; Mismatches 146; Indels 119; Gaps 18;

QY 541 VECRVQLPLRE-YVNAR-HCLP-----CHPEQC 568
DB 354 IEKQESYLLAFEHVINYKHNIPFWSKLMKVADLRMIGAYHASFRLHMKVCEPTELS 413
QY 569 PQNGSVTCGPBACDCAVCAHYKDPFCVAPCPQYIKANSKFGITELPDESGACQCP 628
DB 414 PGE-----VGP--DHCKMCAHFIDGPHCVKACPAGVLGENDTL-VMKYADANAVCOLCHP 465
QY 629 NCTHSCVDLDDGCPAERASPLTSIVSAV-V-GILLVVVLGVVFGILIKRROOKIRKTYM 687
DB 466 NCTRGCKGPLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRR-HIVKRTL 521
QY 688 RLLQETELVEPLTSGAMPNQAQRILKETELRKVKVLGSGAFGTVYKGIWIPGENVK 747
DB 522 RLLQERELVEPLTSGEAPNQAHLRIKETEFKVKVVLGFGAFGTVYKGLIPEGEKVT 581
QY 748 IPVAIKVLRENTSPKANKEILDEAYVMAGVSGPYVSRLLGICLTSTVOLVTLQMPYGC 807
DB 582 IPVAIKELRENTSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQITQMPYGC 641
QY 808 DHVRENRLGSGDILLNWCQIAKGMVSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLA 867
DB 642 DYIREHKDNIQSOLLNWCQIAKGMVSYLEDVRLVHRDLAARNVLKTPQHVKITDFGLA 701
QY 868 RLLDIDETEHADGCKVPKIMWMALESILRRTHQSDVWSYGVTVWELMTFGAKPYDGP 927
DB 702 KQLGADEKEYHAEGKVPKIMWMALESILRIYTHQSDVWSYGVTVWELMTFGSKPYDGP 761
QY 928 AREIPDLLEKGRLLPQPPICITIDVYIMVWKMIIDSECRPRFELVSEFSRMARDPQ 987
DB 762 ASEISSVLEKGRLLPQPPICITIDVYIMVWKMSGADSRPKRELIAEFSKWARDPPRYL 821
QY 988 VIQ-NEDLGPASPLSTFYRSLLEDMDGLVDABEYLVPOQGFPCPDAPGAGGMVHR 1046
DB 822 VIQDERMHLPSPTSKFYRTLMEEDMEDI VDAEYLVPHQGF-----866
QY 1047 HRSSSTRSGGDLTLGLPSEEAERPSPLAPSEGAGSDVFDGLMGAKGLQSLPTHP 1106
DB 867 -NSPST-----SRTPLLSLSATN-----NSATKCIDRNGGH--898
QY 1107 SPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYVNOQVVRPQPSREGPLPAAPAGAT 1166
DB 899 -----PVREDGFL-----PAPEYVNOQ--LMPKKPSTAMVQNIYNI 936

```

```

QY 1167 -LERAKTLSPGKNGVVKVFAFGGAVENPEYL 1197
DB 937 AISKLPMDSRYNQ-----SHSTAVDNPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Ta34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=28971102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:285-278(1987).
DR EMBL; X06943; CAA30024.1; --
DR HSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DC8CCA0F8AF4 CRC64;

Query Match 23.7%; Score 1615; DB 15; Length 545;
Best Local Similarity 54.3%; Pred. No. 1.5e-113;
Matches 341; Conservative 70; Mismatches 125; Indels 92; Gaps 15;

QY 578 GPEADQCVACAHYKDPFCVAPCPQYIKANSKFGITELPDESGACQCPINCHTSCVDL 637
DB 1 GP--DHCKMCAHFIDGPHCVKACPAGVLGENDTL-VMKYADANAVCOLCHPCTRGCKGP 57
QY 638 DRKGCPEAERASPLTSIVSAV-V-GILLVVVLGVVFGILIKRROOKIRKTYMRLLOETEL 696
DB 58 GLEGGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRR-HIVKRTLRELLQEREL 113
QY 697 VEPLTPSGAMPNQAQRILKETELRKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVL 756
DB 114 VEPLTPSGEAPNQAHLRIKETEFKVKVVLGFGAFGTVYKGLIPEGEKVTIPVAIKEL 173
QY 757 ENTSPKANKEILDEAYVMAGVSGPYVSRLLGICLTSTVOLVTLQMPYGCILLDHVREN 816
DB 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQITQMPYGCILLDVIHDKN 233
QY 817 LGSQDLLNWCQIAKGMVSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDDETE 876
DB 234 IGSQDLLNWCQIAKGMVSYLEDVRLVHRDLAARNVLKTPQDVKITDFGLAKOLGDAKE 293
QY 877 YHADGCKVPKIMWMALESILRRTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLE 936
DB 294 YHAEKGKVPKIMWMALESILRIYTHQSDVWSYGVTVWELMTFGSKPYDGI PASEISSVLE 353
QY 937 KGERLPQPPICITIDVYIMVWKMIIDSECRPRFELVSEFSRMARDPQRFVVIQ-NEDLG 995
DB 354 KGERLPQPPICITIDVYIMVWKMSGADSRPKRELIAEFSKWARDPPRYLVYIQGDERMH 413
QY 996 PASPLDSTFYRSLLEDMDGLVDABEYLVPOQGFPCPDAPGAGGMVHRHRSSTRSG 1055
DB 414 LPSPDTSKFYRTLMEEDMEDI VDAEYLVPHQGF-----NSPST---454

```

QY 1056 GGDLTGLPSEEEAPRSL-----APSEGAGSDVFDGLMGCAAGLQSLPHTDPSPLQ 1110
Db 455 -----SRTPLLSLSATSNNATNCIDRNG-----H----- 481
QY 1111 RYSEDTFVLPSTGTVAPLTCSPQPEYVQPDVRPQPSREGPLPAARPAAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVNO--LMPKBPSTAMVQTOIYNYISLTAISK 523
QY 1170 AKTLPCKGKGVKDVFAFGGAVENPEYL 1197
Db 524 LPMDSRYON-----SHSTAVDNPEYL 544

RESULT 15

Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
DE EGFR.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maibie N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maibie N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egr transcripts encoding truncated receptor isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Iehii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Azawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barén G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bona M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -;
DR EMBL; AF275366; AAG28047.1; -;

DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -;
DR EMBL; AK004883; BAB23641.1; -;
DR EMBL; AK004911; BAB23662.1; -;
DR MGD; MGI:95294; Egr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recept_L_domain; 2.
DR SMART; SM00261; FU; 3.
DR Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6834063B1BC928CB CRC64;

Query Match 21.7%; Score 1478.5; DB 11; Length 655;
Best Local Similarity 44.2%; Pred. No. 4.3e-103;
Matches 283; Conservative 94; Mismatches 238; Indels 25; Gaps 9;
QY 11 LLLALLPPGAA--STOVCTGDMKRLPASPETHDMLRLHYQCQVVOGNLEITYLPTN 68
Db 14 LLTALCAAGALBEKKVCOQTSNRLTQGTGFEDHFLSLQRMYNNEVVLGNLEITYVQRN 73
QY 69 ASLSFLQDIOEVQGYVLIHNVQVPLQRLRIVRGTLQFEDNYVALAVLQNGDPLNNTTP 128
Db 74 YDLSFLKTIQEVAGYVLIHNVQVPLQRLRIVRGTLQFEDHFLSLQRMYNNEVVLGNLEITYVQRN 124
QY 129 VTGASPGGLRELQRLSITELIKGVLQIORNPOLCYQDTILWKDI----EKKNNOLALTLI 184
Db 125 -YGTNRGTGLRELPMRLQELIIGAVRFSNPIICNMDTIQWRDVIQVNFMSNMDL--- 180
QY 185 DTNRSRACHPCSPMGKSGRCWGESSEDCQSLRTVTCAGGCA-RCKGPLPTDCCHCEQCAAG 243
Db 181 -QSHPPSCPKCDPSCPNCSGCGGEEENCKLTKIICAQCSHRGRSPSCDCHNQCAAG 239
QY 244 CTGPKSHDCLACLHFNHSGICELHCPALVTYNTDTTFESMPNPEGRYTFGASCVTACPNY 303
Db 240 CTGPRESDECLVCQKQFDEATCKDTCPLMLNPTTYQMDVNPQEGYSGFATCVKKCPRY 299
QY 304 LSTDVSGCTLVCLPHNQEVTAEDGTORCEKSKPCARVCYGLGQYIKANSKFIGITELE 363
Db 300 VVTDHSGSVACPDYEV-EEDGIRKCKCDGPRKVCNGIGIGEFK-DTSLSNATNIK 357
QY 364 -FAGCKKIFGSLAFPESEFDGDPASNTAPLQEQLOVFTLEETLYLYISAWPDSLPDL 422
Db 358 HFKYCTAISGDLHLPLVAFKGSFTRTPPLDPRELEILTKVKEITGFLLIQAWPDNWDL 417
QY 423 SVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRSLRELGLALIHHTHLCFVHTVPW 482
Db 418 HAFENLEIIRGRTKQHGQFSLAVVGLNITSLRLSLKEISDGDVILSGNRNLCYANTINW 477
QY 483 DQLPRPHQALLHTANRPEDECEVGEGLACHOLCARGHCGWPGPTQCVCNCSQFLRGQECVE 542
Db 478 KKLFGTPNQTKTKMNNRAEKDCAVNHVNCPLCSSECGCWGPEPRDCVSCQNVSRGECVE 537
QY 543 ECRVLOGLPREYVNRARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPFCVARCPQ 602
Db 538 KCNILEGEPRFVENSECICHPCECLPQAMNITTCRGPDCNICOAHYIDGPHCVKTCQA 597
QY 603 YIKANSKFITGLPDEEGACQCPINCTHSCVDLDDKGC 642
Db 598 GIMGENNTL-VMKYADANNVCHLCHANCTYGCAGPGLQGC 636

Search completed: July 22, 2003, 09:01:33
Job time : 53.3575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds
(without alignments)
5088.033 Million cell updates/aec

Title: SEQ4-579-593-12

Perfect score: 6803

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6651	97.8	1255	1 ERB2 HUMAN	P04626 homo sapien
2	5862	86.2	1257	1 ERB2 RAT	P06494 rattus norv
3	5946.5	85.9	1254	1 ERB2 MESA	O60553 mesocricetu
4	3114	45.8	1210	1 EGFR HUMAN	P00533 homo sapien
5	3090	45.4	1210	1 EGFR MOUSE	Q01279 mus musculu
6	2956.5	43.5	1308	1 ERB4 HUMAN	Q15303 homo sapien
7	2944	43.3	1308	1 ERB4 RAT	Q62956 rattus norv
8	2657.5	39.1	1167	1 XMRK XIPMA	P13388 xiphophorus
9	2397.5	35.2	1342	1 ERB3 HUMAN	P21860 homo sapien
10	2327.5	34.2	1339	1 ERB3 RAT	Q62799 rattus norv
11	1947	28.6	1426	1 EGFR DROME	P04412 drosophila
12	1719.5	25.3	634	1 ERBB ALV	P00534 avian leuko
13	1688	24.8	604	1 ERBB AVIER	P00535 avian eryth
14	1600	23.5	540	1 ERBB AVIEU	P11273 avian eryth
15	1565	23.0	703	1 EGFR CHICK	P13387 gallus gall
16	1294	19.0	1323	1 LT23 CAEEL	P24348 caenorhabdi
17	1142.5	16.8	245	1 ERB2 MOUSE	P70424 mus musculu
18	723	10.6	1363	1 ILPR BRALA	O02466 branchiosto
19	703	10.3	1300	1 IRR MOUSE	Q9wtc14 mus musculu
20	697	10.2	1607	1 MIPR LYMT	Q25410 lymnaea eta
21	692	10.2	1297	1 IRR HUMAN	P14616 homo sapien
22	691	10.2	1383	1 INSR RAT	P15127 rattus norv
23	686.5	10.1	1300	1 IRR CAVPO	P06213 homo sapien
24	684	10.1	1382	1 INSR HUMAN	P15208 mus musculu
25	682.5	10.0	1372	1 INSR MOUSE	Q25197 hydra atten
26	678	10.0	1477	1 HTK7 HYDAT	P08069 homo sapien
27	641	9.4	1367	1 IGIR HUMAN	O60751 mus musculu
28	623	9.2	1373	1 IGIR MOUSE	P09208 drosophila
29	621	9.1	2146	1 INSR DROME	P24062 rattus norv
30	619.5	9.1	1370	1 IGIR RAT	Q93105 aedes aegypt
31	619	9.1	1390	1 INSR AEDAE	P54760 homo sapien
32	600	8.8	987	1 EPB4 HUMAN	P09759 rattus norv
33	599.5	8.8	984	1 EPB1 RAT	

34 593.5 8.7 984 1 EPB1 HUMAN
35 592 8.7 902 1 EPBB XENLA
36 583.5 8.6 985 1 EPBA_XENLA
37 581.5 8.5 984 1 EPB1_CHICK
38 580 8.5 987 1 EPBA_MOUSE
39 578.5 8.5 1053 1 FAKT_CHICK
40 578 8.5 1114 1 RET_HUMAN
41 575.5 8.5 977 1 EPA2 MOUSE
42 569.5 8.4 976 1 EPA2 HUMAN
43 569 8.4 1068 1 FAK1_XENLA
44 565.5 8.3 757 1 HT16_HYDAT
45 565 8.3 1052 1 FAK1_MOUSE

ALIGNMENTS

RESULT 1
ERB2_HUMAN
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell surface receptor HER2) (MLN 19).
DE ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutharia; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N., Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erbB-2 gene to epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A., McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J., Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal growth factor-receptor gene and is amplified in a human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-specific competition hybridization.";
RL Genomics 15:426-429(1993).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIREGULIN.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC -1- RESIDUES (BY SIMILARITY).
 CC -1- POLYMORPHISM: THERE ARE FOURS ALLELES DUE TO THE VARIATIONS IN
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M11767; AAA35808.1; -
 CC EMBL; M11761; AAA35808.1; JOINED.
 CC EMBL; M11762; AAA35808.1; JOINED.
 CC EMBL; M11763; AAA35808.1; JOINED.
 CC EMBL; M11764; AAA35808.1; JOINED.
 CC EMBL; M11765; AAA35808.1; JOINED.
 CC EMBL; M11766; AAA35808.1; JOINED.
 CC EMBL; M11730; AAA35978.1; -
 CC EMBL; M12036; AAA35978.1; -
 CC EMBL; X03363; CAA27060.1; -
 CC PIR; A25491; A25491.
 CC PIR; A25771; A24571.
 CC HSP; P11362; IFGK.
 CC Genew; HGNC:3430; ERBB2.
 CC MIM; 164870; -
 CC InterPro; IPR000494; EGFR_L domain.
 CC InterPro; IPR000719; Euk_pkinase.
 CC InterPro; IPR002174; Furin-like.
 CC InterPro; IPR001245; Tyr_pkinase.
 CC InterPro; IPR004019; YLP_motif.
 CC Pfam; PF00069; pkinase; 1.
 CC Pfam; PF00757; Furin-like; 1.
 CC Pfam; PF01030; Recep_L_domain; 2.
 CC Pfam; PF02757; YLP_2_domain; 1.
 CC ProDom; PD000001; Euk_pkinase; 1.
 CC SMART; SM00261; FU; 3.
 CC SMART; SM00219; TyrKc; 1.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 CC PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
 CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 CC Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 CC Polymorphism.
 CC SIGNAL 1 21
 CC CHAIN 22 1255
 CC DOMAIN 22 652
 CC TRANSMEM 653 675
 CC DOMAIN 676 1255
 CC NP BIND 726 734
 CC BINDING 753 753
 CC ACT SITE 845 845
 CC DISULFID 195 204
 CC DISULFID 199 212
 CC DISULFID 220 227
 CC DISULFID 224 235
 CC DISULFID 236 244
 CC DISULFID 240 252
 CC DISULFID 255 264
 CC DISULFID 268 295
 CC DISULFID 299 311
 CC DISULFID 315 331
 CC DISULFID 334 338
 CC DISULFID 511 520
 CC DISULFID 515 528
 CC DISULFID 531 540
 CC DISULFID 544 560

FT	DISULFID	563	576	BY SIMILARITY.
FT	DISULFID	567	584	BY SIMILARITY.
FT	DISULFID	587	596	BY SIMILARITY.
FT	DISULFID	600	623	BY SIMILARITY.
FT	DISULFID	626	634	BY SIMILARITY.
FT	DISULFID	630	642	BY SIMILARITY.
FT	MOD RES	1139	1139	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD RES	1248	1248	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	68	68	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	124	124	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	187	187	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	259	259	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	530	530	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	571	571	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	629	629	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARIANT	654	654	I -> V.
FT	VARIANT	655	655	/FTid=VAR_004077.
FT	VARIANT	655	655	I -> V.
FT	CONFLICT	1170	1170	/FTid=VAR_004078.
FT	CONFLICT	1255	1255	P -> A (IN REF. 2).
FT	SEQUENCE	1255	1255	AA; 137909 MW; 39E9DFA04DCF962 CRC64;
Qy	Query Match			97.8%; Score 6651; DB 1; Length 1255;
Db	Best Local Similarity			97.6%; Pred. No. 0;
Qy	Matches 1225; Conservative			10; Mismatches 20; Indels 0; Gaps 0;
Qy	1	MELAAALCRWGLLLALLPPGAAS	TOVCTGTDKMLRLPASPETHLDMLRHL	YQCCVQGNL 60
Db	1	MELAAALCRWGLLLALLPPGAAS	TOVCTGTDKMLRLPASPETHLDMLRHL	YQCCVQGNL 60
Qy	61	ELTYLPTNASLSFLQDIOEQVGY	LIQHNQVQVPLQRLIRVRGTQ	LFEDNYALVLDNG 120
Db	61	ELTYLPTNASLSFLQDIOEQVGY	LIQHNQVQVPLQRLIRVRGTQ	LFEDNYALVLDNG 120
Qy	121	DPLNNTTPVTGASPGGLREQLR	SLTEILKGGVLIQHNQVQVPLQ	RLIRVRGTQLFEDNYALVLDNG 180
Db	121	DPLNNTTPVTGASPGGLREQLR	SLTEILKGGVLIQHNQVQVPLQ	RLIRVRGTQLFEDNYALVLDNG 180
Qy	181	LTLIDTNRSRACHPCSPMCKGSR	CGWSESSDCQSLTRTVCA	GCGCARCKGPLPTCCHEQC 240
Db	181	LTLIDTNRSRACHPCSPMCKGSR	CGWSESSDCQSLTRTVCA	GCGCARCKGPLPTCCHEQC 240
Qy	241	AAGCTGPKHSDCLACLFHNSG	ICELHCPALVTNTDTFESMPN	PEGRYTFGASCVTACP 300
Db	241	AAGCTGPKHSDCLACLFHNSG	ICELHCPALVTNTDTFESMPN	PEGRYTFGASCVTACP 300
Qy	301	YNYLSTDVSGSCTLVCP	PLHNOEVTAEADGTQRCCK	SKPCARVCYGLGMOYIKANSKF 360
Db	301	YNYLSTDVSGSCTLVCP	PLHNOEVTAEADGTQRCCK	SKPCARVCYGLGMEHLREVR 360
Qy	361	ELFEPAGCKIFGSLAFIPES	FDGDPASNTAPLQEQLOV	PETLEITCYLYISAWPDSL 420
Db	361	IQEFAGCKKIFGSLAFIPES	FDGDPASNTAPLQEQLOV	PETLEITCYLYISAWPDSL 420
Qy	421	DLVSFQNLQVIRGRILHNGAY	SLTLOGLGISWLGRLSRL	ELGSLALHNNTHLCFVHTV 480
Db	421	DLVSFQNLQVIRGRILHNGAY	SLTLOGLGISWLGRLSRL	ELGSLALHNNTHLCFVHTV 480
Qy	481	PWDQLFRNPHOALLHTANR	PEDECVGEGGLACHQI	LCARGHCHMGPGPTQCVCNSQ 540
Db	481	PWDQLFRNPHOALLHTANR	PEDECVGEGGLACHQI	LCARGHCHMGPGPTQCVCNSQ 540
Qy	541	VEECRVLOGLPREYVNAHCL	PCHPCEOPONGSVTCF	PEADOCVACAHYKDPFCV 600
Db	541	VEECRVLOGLPREYVNAHCL	PCHPCEOPONGSVTCF	PEADOCVACAHYKDPFCV 600
Qy	601	PQVIKANSKFIGITELP	DEEGACQPCPINCTHSC	VDLDDKCPAEQASPLT 660
Db	601	PSGVKPDLSYMPINWK	PFDEEGACQPCPINCTH	SCVDLDDKCPAEQASPLT 660
Qy	661	ILLVVLGVVFGILIKRRQ	QKIRKYTWRRLLQETEL	VEPLTPSGAMPNQMRILK 720
Db	661	ILLVVLGVVFGILIKRRQ	QKIRKYTWRRLLQETEL	VEPLTPSGAMPNQMRILK 720


```

Oy 721 RKVKVLSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVAGVGP 780
Db 721 RKVKVLSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVAGVGP 780
Oy 781 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRENRGRGLSQDLLNWCQIAKGM5YLEDVR 840
Db 781 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRENRGRGLSQDLLNWCQIAKGM5YLEDVR 840
Oy 841 LVHRDLARNVLYKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRRFT 900
Db 841 LVHRDLARNVLYKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRRFT 900
Oy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKCWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKCWM 960
Oy 961 IDSECRPRFRELSEFMRMDPQRFVVIQNEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
Db 961 IDSECRPRFRELSEFMRMDPQRFVVIQNEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
Oy 1021 EEVLVPOOGFCFDPAPGAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSPAPSEG 1080
Db 1021 EEVLVPOOGFCFDPAPGAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSPAPSEG 1080
Oy 1081 AGSDVFDGLGMAKGLQSLPTHDPSPLORYSEDPTVLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAKGLQSLPTHDPSPLORYSEDPTVLPSETDGYVAPLTCSPQPEYV 1140
Oy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGNVGVKOVFAFGGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGNVGVKOVFAFGGAVENPEYLTPO 1200
Oy 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPPSTPKGTPPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPPSTPKGTPPTAENPEYGLDVPV 1255

RESULT 2
ID ERB2_RAT STANDARD; PRT; 1257 AA.
AC F06494;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (epidermal growth factor
DE receptor-related protein).
GN ERBB2 OR NEU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
RX MEDLINE=86118662; PubMed=3945311;
RT "The neu oncogene encodes an epidermal growth factor receptor-related
RT protein.";
RL Nature 319:226-230(1986).
RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Lofth F.J., Doak D.G., Mulvey D.,

```

```

RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-
RL oncogenic and oncogenic forms of the neu protein.";
EMBO J. 11:43-48(1992).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTOUGHOUT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03362; CAA27059.1; ALT_INIT.
DR PIR; A24562; TVRTNU.
DR HSP; P11362; IFKG.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recept_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 677 POTENTIAL.
FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 159 369 CYS-RICH.
FT DOMAIN 473 646 CYS-RICH.
FT DOMAIN 722 989 PROTEIN KINASE.
FT NP_BIND 728 736 ATP (BY SIMILARITY).
FT BINDING 755 755 ATP (BY SIMILARITY).
FT ACT_SITE 847 847 BY SIMILARITY.
FT DISULFID 196 205 BY SIMILARITY.
FT DISULFID 200 213 BY SIMILARITY.
FT DISULFID 221 228 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 312 BY SIMILARITY.
FT DISULFID 316 332 BY SIMILARITY.
FT DISULFID 335 339 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 578 BY SIMILARITY.

```

```
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 6129264583011402 CRC64;

Query Match 86.2%; Score 5862; DB 1; Length 1257;
Best Local Similarity 86.1%; Pred. No. 9.4e-308;
Matches 1082; Conservative 56; Mismatches 117; Indels 2; Gaps 2;

QY 1 MELAAACRWGLLALLPAGASTOVCTGTDKMLRLPASPETHLDMLRHLVGGCQVQGNL 60
DB 1 MELAAACRWGLLALLPAGTAGTQCTGTDKMLRLPASPETHLDMLRHLVGGCQVQGNL 60
QY 61 ELTYLPTNASLFLQDIQOEVGYVLIAHNOVROVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
DB 61 ELTYVPAASLFLQDIQOEVGYVLIAHNOVQVPLQRLRIVRGTOQLFEDKIALAVLDNR 120
QY 121 DPLNNTTPVT-GASPGGLRELQLRLSTEILKGGVLIORNPOLCYQDTILMKDIFPHKNQL 179
DB 121 DPQDNVAASTPGRTPGELRELQLRLSTEILKGGVLIORNPOLCYQDMVLMKDFVRKNQL 180
QY 180 ALTLIDNRSRACHPCSPMKGSRGWSSESSEDCSLRTVCAGCARCKGRLPDTDCHEQ 239
DB 181 APVDIDNRSRACHPCSPMKGSRGWSSESSEDCSLRTVCAGCARCKGRLPDTDCHEQ 240
QY 240 CAAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTAC 299
DB 241 CAAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTTC 300
QY 300 PYNVLTSDVSGTCLVCPHNOEVTAEQGTQRCCKSPCARVCYGLGQYIKANSKFIGI 359
DB 301 PYNVLTSDVSGTCLVCPHNOEVTAEQGTQRCCKSPCARVCYGLGMEHLRGARITSD 360
QY 360 TELEFACKTIFGLSLFESFDCGDPASNTAPLOEQLOVPETLEEITGYLIYSAWDSL 419
DB 361 NVOEFDGCKKIFGLSLFESFDCGDPSSGIAPLPEQLOVPETLEEITGYLIYSAWPSL 420
QY 420 PDLVSFQNLQVIRGILHNGAYSILTQGLGTSWGLSLRSLRSLGSLALIHNTLCFVHT 479
DB 421 RDLVSFQNLRIIRGILHDGAYSILTQGLGHSGLSLRSLRSLGSLALIHNAHLCFVHT 480
QY 480 VPMQDLFRNPHQALLHTANRPEDE-CVGEGLACHQLCARGHCWGPGTQCVCNCSQFLRGQ 538
DB 481 VPMQDLFRNPHQALLHSGNRPEDDGLVSSGLVNCNLSLCAHGHCHWGPGTQCVCNCSHFLRGQ 540
QY 539 ECVEECRVQLQPLREYNARHCLPCHPEQCPQNGSVTCFCGPEADQCVACAHYKDPFCVA 598
DB 541 ECVEECRVWGLPREYVSDRKLCPCHPEQCPQNSSETCFGEADQCAAHYKDSSCVA 600
QY 599 RCPQYIKANSKFIGITELPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIYSAV 658
DB 601 RCPSGVKPDLISYMPIKYDEEGICQPCINCTHSCVDLDERGCPAQASPLTSIYSAV 660
QY 659 VGILLVVVLGVFGILIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKET 718
DB 661 VGILLVVVLGVFGILIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKET 720
QY 719 ELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAVMAGVG 778
DB 721 ELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAVMAGVG 780
QY 779 SPVVSRLGLICLTSTVOLVTLQMPYGLDLHVRNRRGLSGQDLLNMCQIAKMSYLE 838
```

```
DB 781 SPVVSRLGLICLTSTVOLVTLQMPYGLDLHVRHRRGLSGQDLLNMCQIAKMSYLE 840
QY 839 VRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGCKVPIKMALESILRRR 898
DB 841 VRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGCKVPIKMALESILRRR 900
QY 899 FTHQSDVMSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPOPPICITIDVYMIWVC 958
DB 901 FTHQSDVMSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPOPPICITIDVYMIWVC 960
QY 959 WMIDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDEDDMGDLV 1018
DB 961 WMIDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDEDDMGDLV 1020
QY 1019 DABEYLVPQGGFFCPDPAPGAGGMVHRHRSSTSRGSGDLTLGLSESEEEAPRSLAPS 1078
DB 1021 DABEYLVPQGGFFCPDPPTGCTAHRHRSSTSRGSGDLTLGLSESEEGPPRSLAPS 1080
QY 1079 EGAGSDVDFDGLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCPOPE 1138
DB 1081 EGAGSDVDFDGLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCPOPE 1140
QY 1139 YVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVKDVFAGGAVENPEYLT 1198
DB 1141 YVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVKDVFAGGAVENPEYLT 1200
QY 1199 POGGAAPQPPHPPAFSPAFDNLVYWDODPERRGAPPSTFKGTPTAENPEYLGDLV 1255
DB 1201 PREGTAPPPHPPAFSPAFDNLVYWDQNSSEQGPSPNFECTPTAENPEYLGDLV 1257

RESULT 3
ID ERB2 MESAU STANDARD; PRT; 1254 AA.
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DB (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```



```

Oy 1141 NQPDVVRQPPSPREGPLPAARACATLRLAKTLSPGKGVVQVFAFGAVENPEYLTPQ 1200
Db 1141 NQEVRRQPPPLTEGCPPLPPVRPAGATLRLPKTLSPGKGVVQVFTFGGAVENPEYLVR 1200
Oy 1201 GGAAPQHPHPPAFSPAFDNLYYWDQPPRGAPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGSASQPH-PPALCPAFDNLYYWDQPPSERGSPNTPEGTPTAENPEYGLDVPV 1254

RESULT 4
EGFR HUMAN
ID EGFR HUMAN STANDARD; PRT; 1210 AA.
AC P00533; P06268; Q14225; Q9UMD8; Q9UMG5; Q92795; Q00732;
AC Q00688; Q9B2S2; Q9H2C9; Q9GZX1; Q9H3C9;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUN-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
DE protein-tyrosine kinase ErbB-1).
GN EGFR OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RA "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL Nature 309:418-425(1984).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta;
RA Ilekis J.V., Stark B.C., Scoccia B.;
RA MEDLINE=95382957; PubMed=7654368;
RT "Possible role of variant RNA transcripts in the regulation of
RT epidermal growth factor receptor expression in human placenta.";
RL Mol. Reprod. Dev. 41:149-156(1995).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta;
RA MEDLINE=97078686; PubMed=8918811;
RA Reiter J.L., Maible N.J.;
RT "A 1.8 kb alternative transcript from the human epidermal growth
RT factor receptor gene encodes a truncated form of the receptor.";
RL Nucleic Acids Res. 24:4050-4056(1996).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta;
RA MEDLINE=97256547; PubMed=9103388;
RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
RT "Expression of a truncated epidermal growth factor receptor-like
RT protein (TEGFR) in ovarian cancer.";
RL Gynecol. Oncol. 65:36-41(1997).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RX TISSUE=Placenta;
RA MEDLINE=21100872; PubMed=11161793;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative EGFR transcripts encoding truncated receptor
RT isoforms.";
RL Genomics 71:1-20(2001).
RN [6]
RP SEQUENCE OF 575-687 FROM N.A.
RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,

```

```

RA Maible N.J.;
RT "Human and mouse alternative EGFR transcripts encoding only the
RT extracellular domain of the receptor.";
RL Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; PubMed=6326261;
RA Lin C.R., Chen W.S., Krueger W., Stolarek L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT "Expression cloning of human EGF receptor complementary DNA: gene
RT amplification and three related messenger RNA products in A431
RT cells.";
RL Science 224:843-848(1984).
RN [8]
RP SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245835; PubMed=6330563;
RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RT "Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAs overproduced in A431 carcinoma cells.";
RL Nature 309:806-810(1984).
RN [9]
RP SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; PubMed=6093780;
RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
RA O'Malley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RT receptor cDNA from human A431 carcinoma cells.";
RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN [10]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=88217333; PubMed=3329716;
RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.D.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription.";
RL Oncogene Res. 1:375-396(1987).
RN [11]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; PubMed=1988448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis.";
RL J. Biol. Chem. 266:1746-1753(1991).
RN [12]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; PubMed=2991899;
RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN [13]
RP SEQUENCE OF 540.
RA Kohda D.;
RT Submitted (SEP-1997) to the SWISS-PROT data bank.
RN [14]
RP RECEPTOR ACTIVITY.
RX MEDLINE=84191554; PubMed=6325948;
RA Mroczkowski B., Morig G., Cohen S.;
RT "ATP-stimulated interaction between epidermal growth factor receptor
RT and supercoiled DNA.";
RL Nature 309:270-273(1984).
RN [15]
RP PHOSPHORYLATION.
RX MEDLINE=89278137; PubMed=2543678;
RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
RT "All autophosphorylation sites of epidermal growth factor (EGF)
RT receptor and HER2/neu are located in their carboxyl-terminal tails.
RT Identification of a novel site in EGF receptor.";
RL J. Biol. Chem. 264:10667-10671(1989).
RN [16]

```


CC -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS. NRG-
 CC 2. NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
 CC NTA. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC -!- NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphatase.
 CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLAVABLE AND
 CC NONCLAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
 CC LUNG, SALIVARY GLAND, AND PANCREAS.
 CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L07868; AAB59446.1; -
 CC HSPG: P11362; LFQK
 CC Genew: HGNC:3432; ERBB4.
 CC MIM: 600543; -
 CC InterPro: IPR000494; EGFR_L_domain.
 CC InterPro: IPR000719; Euk_pkinase.
 CC InterPro: IPR002174; Furin-like.
 CC InterPro: IPR001245; Tyr_pkinase.
 CC InterPro: IPR004019; YLP_motif.
 CC Pfam: PF00069; pkinase; 1.
 CC Pfam: PF00757; Furin-like; 1.
 CC Pfam: PF01030; Recep_L_domain; 2.
 CC Pfam: PF02757; YLP; 2.
 CC ProDom: PD000001; Euk_pkinase; 1.
 CC SMART: SM00261; FU; 4.
 CC SMART: SM00219; Tyrc; 1.
 CC PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 CC PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 CC PROSITE: PS00011; PROTEIN KINASE DOM; 1.
 CC Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;
 CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 CC Alternative splicing.
 CC SIGNAL 1 25
 CC CHAIN 26 1308
 CC DOMAIN 26 651
 CC TRANSMEM 652 675
 CC DOMAIN 676 1308
 CC DOMAIN 186 334
 CC DOMAIN 496 633
 CC DOMAIN 718 985
 CC NP BIND 724 732
 CC BINDING 751 751
 CC ACT SITE 843 843
 CC DISULFID 189 197
 CC DISULFID 193 205
 CC DISULFID 213 221
 CC DISULFID 217 229
 CC DISULFID 230 238
 CC DISULFID 234 246
 CC DISULFID 249 258
 CC DISULFID 262 289
 CC DISULFID 293 304

FT DISULFID 308 323 BY SIMILARITY.
 FT DISULFID 326 330 BY SIMILARITY.
 FT DISULFID 503 512 BY SIMILARITY.
 FT DISULFID 507 520 BY SIMILARITY.
 FT DISULFID 523 532 BY SIMILARITY.
 FT DISULFID 536 552 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 559 577 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT DISULFID 593 614 BY SIMILARITY.
 FT DISULFID 617 625 BY SIMILARITY.
 FT DISULFID 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPLIC 626 648 NGPTSDCIYPPMTGHSTLPQHA -> IGSSIEDICIGLMD
 FT (IN ISOFORM JM-B).
 SQ SEQUENCE 1308 AA; 146807 MW; 5E4AE80985D88761 CRC64;
 Query Match 43.5%; Score 2956.5; DB 1; Length 1308;
 Best Local Similarity 45.2%; Pred. No. 1.4e-151;
 Matches 609; Conservative 181; Mismatches 383; Indels 173; Gaps 28;
 Qy 9 WGLILALLPPGAA-----STOVCTGTDMLRLPASPETHLDMLRLHLYOCQVVOGNLEITY 64
 Db 8 WVVVSLVAAGTVQPSDSQSVCACTENKLSLSLEQOYRALRYKYENCEVVMGNLEITS 67
 Qy 65 LPTNASLFLQDIOEVQGYVLIHNNQVQVPLQRLIRVGRQQLFEDNYVALVONGDPLN 124
 Db 68 IEHRDLFLRSRVREVTGYVLVALNQPRYLPLENLRIIRGTLKYEDRYALAIFLNRYKDG 127
 Qy 125 NTPVTVGASPGGLRELQRLRSITELKGGVLIQRPOLCYODTILKWKDIFHKNQNALTLI 184
 Db 128 NF-----GLOELGLKNLTELNGGVYVDQNKFLCYADTIIHWQDILVRNPWSNLTFLV 178
 Qy 185 DTRNSRACHPCSPMKSGRCWGESSEDCQSLTRTVACAGC-ARCKGPLPTDCCHQCAAG 243
 Db 179 STNGSSGGRCHKSCCTG-RWGPPTENHCQTLTRTVCAEQCDGRCYGPVSDCHRECAGG 237
 Qy 244 CTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPVNY 303
 Db 238 CSGPKDTCDFACMNFNDSGACVTCQTPOTFVYNPTTFOLEHNFNAKYTYGAFVCKKCHNF 297
 Qy 304 LSTDVGSCTLVCPLNQOEVTAEQDTCRCEKSKPCARVCYGLGMOYIKANSKFICITILE 363
 Db 298 V-VDSSCVCRACPSKMEV-EENGIRKMKCPCTDICPKACDGIPTGSLMSAQTVDSSNDK 355
 Qy 364 FAGCKTIFGSLAFPEFDGDPASNTAPLOEQVFEETLEETIGYLYISAWPSLPDLS 423
 Db 356 FINCTKINGNLIIFLVGTGIHGDYPYNAIEAIDPEKLNVTREITVREITGFLNQSWPNMTDFS 415
 Qy 424 VFONLQVIRGRIHNGAYSILTLOGLGISWLGSLRSLRELGLGSLAIHNNTHLCFVHTVPWD 483
 Db 416 VFSNLVTIGRVLVYSGLSLILKQOGITSLSQFQSLKEISAGNIYITDNLNLCYVHTINWT 475
 Qy 484 QLERNPHQALLHTANREDECVGEGGLACHOLCARGHCWGPCTOCVNCQSLRQECVEE 543
 Db 476 TLFSTINQRIVRIDNRKAENCTAGBMVCNHLCSDDGCMGPGDQCLSCRRFSRGRICIES 535
 Qy 544 CRVLQGLPREYVVARHCLPCHPECPQ-QNGSVTCFGEADQCAAHYKDPFPFVARCPQ 602

```
Db 536 CNLYDGEFREFENGSIQCECDPOCEKMEGDLTCHGPGDNCCKSHFKDGNPCVKECPD 595
Qy 603 YIKANSKFIQITGLPDEGACQPCPINCTHSCVDLDDKGC-----PABQRASPL 651
Db 596 GLOGANSFIKYADPDRE--CHPCHPNCQCCNGPTSHDCIYYPWTHGSHLTPQHAR-TPL 652
Qy 652 TSIVSAVV-GILLVVLGVVFGILIKERQOKIRKYTWRLLOETELVELPLTPSGAMPNOA 710
Db 653 --TAAGVIGLGLVIVGLTFAVTVRRKSIK-KKRALRRL-ETELVELPLTPSGTAPNOA 708
Qy 711 QMRILKETELRKVKVLGSGAGFYVYGIWIPDGENVKIPVAIKVLRENTSPKANKETLDE 770
Db 709 QLRILKETELRKVKVLGSGAGFYVYGIWIPDGENVKIPVAIKVLRENTSPKANKETLDE 768
Qy 771 AYVAGVGSYVRLGLICITSTVQLVQLMPYGLLDHVRNRRGLSGDQLLNCMQIA 830
Db 769 ALIMASMDHPLVRLGLVCLSPITQLVQLMPHGLLEYVHEHKDNIGSQLLNWCVOIA 828
Qy 831 KGSYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKMA 890
Db 829 KGMYLEERLVRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKMA 888
Qy 891 LESILRRRTHQSDVMSGYVYVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTID 950
Db 889 LECIHYRKPTHQSDVMSGYVYVWELMTFGCKPYDGPAREIPDLLEKGERLPPOPICTID 948
Qy 951 VYIMVKWCMIDSECRPRFRELVSERSMARDPQRFVIONED-LGPASPLDSTFYRSL 1009
Db 949 VYIMVKWCMIDSECRPRFRELVSERSMARDPQRFVIONED-LGPASPLDSTFYRSL 1008
Qy 1010 EDDMDGLDVAEYLVPOQGFCCPDAPGAGGMVHRHRSSTRSGGDLTLGLEPSEE 1069
Db 1009 DEEDLEDMDAEYLV-PQAFNIPPP-----IYTSRARIDSNRS-----EIGHSPPPAY 1056
Qy 1070 APRS-----PLAP-SEGAGSDVDFDGLGMAAGKLOS 1100
Db 1057 TWMGNGQFVYRDGGFAAQGVSPYRAPTSITIEAPVAQATAEIPDDSCCNLTLRKPA 1116
Qy 1101 LPHDPSPLORVSEDITVPLS-----ETDGYVAPLTCSPQPEYVQNPQVPPSPR 1153
Db 1117 PHVQEDSSQRYASDPTVFAPERSPRGELDEGYTMPDRKPKQEYLNVE----- 1167
Qy 1154 EGPLPAARPAAGATLERAKTLPCKNGVVKDVFAFGAVENPEVLTQGGNAPOPHPPA- 1212
Db 1168 ENPFVSR-----KNGDLQ-----ALDNPYHNASNG-----PPKAE 1199
Qy 1213 -----FSPAFNLVYWDQDPPPERGA--PPSTF 1237
Db 1200 DEVVNEPLYNTFANTLGRAEYLNKILSMPEKAKKAFDNPVWNHSLPRSTLQHPDYL 1259
Qy 1238 KGTP-----AENPEYL 1249
Db 1260 QEYSTKYFYKONGRIRPIVAENPEYL 1285

RESULT 7
ERR4_RAT
ID ERR4_RAT STANDARD; PRT; 1308 AA.
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
CN ERR4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]_TaxID=10116;
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
```

```
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; AF041838; AAC08899.1;
CC EMBL; U52531; AAC53051.1;
CC HSSP; P11362; IFGK.
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; YLP_pkinase.
CC InterPro; IPR004019; YLP motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC PRINTS; PF02757; YLP; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 4.
CC SMART; SM00219; TyKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
```

FT	DOMAIN	186	334	CYS-RICH.	
FT	DOMAIN	496	633	CYS-RICH.	
FT	DOMAIN	718	985	PROTEIN KINASE.	
FT	NP BIND	724	732	ATP (BY SIMILARITY).	
FT	ACT SITE	751	751	BY SIMILARITY.	
FT	DISULFID	189	197	BY SIMILARITY.	
FT	DISULFID	193	205	BY SIMILARITY.	
FT	DISULFID	213	221	BY SIMILARITY.	
FT	DISULFID	217	229	BY SIMILARITY.	
FT	DISULFID	230	238	BY SIMILARITY.	
FT	DISULFID	234	246	BY SIMILARITY.	
FT	DISULFID	249	258	BY SIMILARITY.	
FT	DISULFID	262	289	BY SIMILARITY.	
FT	DISULFID	293	304	BY SIMILARITY.	
FT	DISULFID	308	323	BY SIMILARITY.	
FT	DISULFID	326	330	BY SIMILARITY.	
FT	DISULFID	503	512	BY SIMILARITY.	
FT	DISULFID	507	520	BY SIMILARITY.	
FT	DISULFID	523	532	BY SIMILARITY.	
FT	DISULFID	536	552	BY SIMILARITY.	
FT	DISULFID	555	569	BY SIMILARITY.	
FT	DISULFID	559	577	BY SIMILARITY.	
FT	DISULFID	580	589	BY SIMILARITY.	
FT	DISULFID	593	614	BY SIMILARITY.	
FT	DISULFID	617	625	BY SIMILARITY.	
FT	DISULFID	621	633	BY SIMILARITY.	
FT	MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY).
FT	MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY).
FT	MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY).
FT	MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC.)	(POTENTIAL).
FT	CARBOHYD	174	174	N-LINKED (GLCNAC.)	(POTENTIAL).
FT	CARBOHYD	253	253	N-LINKED (GLCNAC.)	(POTENTIAL).
FT	CARBOHYD	358	358	N-LINKED (GLCNAC.)	(POTENTIAL).
FT	CARBOHYD	410	410	N-LINKED (GLCNAC.)	(POTENTIAL).
FT	CARBOHYD	473	473	N-LINKED (GLCNAC.)	(POTENTIAL).
FT	CARBOHYD	495	495	N-LINKED (GLCNAC.)	(POTENTIAL).
FT	CARBOHYD	548	548	N-LINKED (GLCNAC.)	(POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC.)	(POTENTIAL).
FT	CARBOHYD	620	620	N-LINKED (GLCNAC.)	(POTENTIAL).
FT	CONFLICT	1062	1062	S -> N (IN REF. 3).	
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	
SEQ	SEQUENCE	1308 AA;	146957 MW;	D944BB096A08B41 CRC64;	
Query Match					
Best Local Similarity					
Matches 606; Conservative 190; Mismatches 391; Indels 160; Gaps 28;					
Qy	1	MELA-ALCRWGLLL--	ALLPPGAAS	TQVCTDMKRLRLPASPETHLDMRLHLYOGCQVQ	57
Db	1	MKLATGLWVNGSLVAARTVQPSASQSVACAGTENKLSLSLDLEQQYRALRKYENCCEVVM	60		
Qy	58	GNLELTYLPNASSLFLODTEQVGYVLIHNOVRQVPLORLRIVRGTOQLFEDNYALAVL	117		
Db	61	GNLEITSIEHNRDLSFLRSIREVTGYVALNQPRLPLENRLIRIRGTLKYEDRYALAIF	120		
Qy	118	DNGDPLNNTTPVTGASFGGLREQLRSITELKGGVLJQRNPQLCYODTILWKDIFHKNV	177		
Db	121	LNTRKQGNF-----	GLQELGKLNLTETLNGGVYVDQNKFLCYADTIHQDIVRNPW	171	
Qy	178	QLALTLTDNRSRACHPCSPMKSGRCWGESSEDSQSLIRTVACAGGC-ARCKGLPTDCC	236		
Db	172	PSNMTLVSTIGSSCGRCHKSCGTG-RCWGPTEHHCQLTRTVCAEQCDGRCYGYVSDCC	230		
Qy	237	HEQCAAGCTGPKHSIDLACHLHFNHSGICEUHCPLAVTYNTDTFSPMPNPEGRYTFGASCV	296		
Db	231	HRECAGCGSGPKDTCDFACMFNDSGACVTCQPTQFVYVNTPTTFQLEHNFNAKYTYGAFV	290		
Qy	297	TACPYNVLSLTDVGSCTLVCPHLNDEVTAEDGTQRCCKSKPCARVCYGLGNQYIKANSKF	356		
Db	291	KKCPHNFV-VDSSECVACRACPSKMEV-EENGIMCKRCPCTDICPKACDGIPTGSLMSAQTV	348		

RESULT 8
XMRK_XIPMA
ID_XMRK_XIPMA
AC_P13388;

1237 FKGTPT-----AENPEYL 1249
1259 LQEVSTKYFYKONGRIRPIVAENPEYL 1285

1199 EDEVNEPLYNTFTNALGNAEYKNSLLSVPEKAKAFDNDPDYWNHSLPRSTRLOHPDY 1258

1213 -----FSPAFLNYWDQDPPERGA--PPST 1236

1168 -ENPFVSRR-----KNGDLQ-----ALONPEYHSASSG-----PPKA 1198

1153 REGPLPAARAGATLEBAKTLSPCKGVKDVFAFGCAVENPEYLTQGGGAPOPHPPA 1212

1116 VPHVEDSSQTSYASADPTVPAPERPNRAELDEEGYMTPMHDKPKOEVLPNVE----- 1167

1100 SLPHDPSLQRYSESDPTVPLPS-----ETDGYVAPLTCSPQPEYVNPQDVPBPSP 1152

1061 GSQFVYQDGGFATQCG--NMPYATATSTIPEAPVA--QGATAEMFDDSCCNGTLRKPV 1115

1040 GGMVHHRHRSSTRSGGDLTLGLEPSEEBAPRSLAPSEAGSDVDFDGLGMAAGLQ 1099

1002 KFFQNLLEDLEDMDABEYLP-QAFNIPPPYITSTRIDSNRSEIGHSPPAYTPMS 1060

1003 TFYRSLLDDMDGLVDABEYLPQCGFFCPDP-----APGA 1039

942 PPCTIDVYIMVKCMWIDASRPKFELAAEFSSRMARDPQRYLVIOGDDRMKLPSPNDS 1001

944 PPCTIDVYIMVKCMWIDSECRPRFELVSFSSRMARDPQRFVVIQNEED-LGPASPLDS 1002

882 MPKWMALCIIHYKFTHQSDVMSYGTIWEMLTFFGCKPYDGIPTREIPDLLEKGERLPQ 941

884 VPKWMALCIIHYKFTHQSDVMSYGTIWEMLTFFGCKPYDGIPTREIPDLLEKGERLPQ 943

822 NWCQVQAKGMYLEERLVRDLAARNVLSKPNHVKITDFGLARLLEGDEKEYNADGGK 881

824 NWCQVQAKGMYLEERLVRDLAARNVLSKPNHVKITDFGLARLLEGDEKEYNADGGK 883

762 NVEFMDEALIMASVDHPHLVRLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNIGSOLL 821

764 NKEILDEAYVMAGVGSPPVSRLLIGICLTSTVQLTQMLPYGCLLDHVRNRRGLGSDLL 823

702 GTAPNQAQLRLKTELKRVKVLGSGAFVYKGIWVPEGETVKIPVAIKILNETTGPKA 761

704 GAMPNQAQMRILKETELKRVKVLGSGAFVYKGIWIPDGENVKIPVAIKVLRNENTSPKA 763

647 HAR-TPL--IAAGVIGGLFVLVIMALTFAVYVRKSIK-KKRALRRFL-ETELVEPTPS 701

645 EQRASPLTSTVSADV-GILLVVLGVVFGILIKRQOKIRKTYMRRLLQETELVEPTPS 703

589 CVEKCPDLQGANSP--IFKYADQDRCHPCPNCTQCGNPGTSHDCIYYPWTGHSTLPQ 646

596 CVARCPQYIKANSKFIGITELPDEBGACQPCPNCTHSCVDLDDKGC-----PA 644

529 GKICIESCNLYDGEFREFENGSIQVECDSCERKMEDGLLTCHGPGPNCTKCSHFKDPGN 588

537 GQECVBECEVLOGLPREYVNNARHCLPCHPECPQ-ONGSVTCFGEADQCACAHYKDPFF 595

469 YHTINMTTLFSTVYQRIVRDNRRAENCTAEGVMVCHLCSNDGCGWGPQDCLSCRRFSR 528

477 VHTVPMDQLPRNPHQALLHTANPEDECEVGEGLACHOLCARHGCHWGPQTCVNCOSFLR 536

409 PNMTDFSVSNLVTIGRVLVSLLLILKQOQITSLQFOSLKEISAGNIYITDINSNLCY 468

417 DSLPDLVSFONLQVIRGRILHNGAYSLTLQGLGISMGLSLRSLRGLSGLALIHNNHLCF 476

349 DSSNIDKFNCTKINGNLFLVGTGIHGDPPYNAIDAIDPEKLNVRFTVREITGFLNIOTWP 408

357 IGITELEFAGCKKIFGSLAPLPESFDGDPASNTAPLOEQLOVFETLEEITGVLYISAMP 416


```

Db      590  IPRCPHGLDGTLL-IWKYADKMGQCPCHQCTQCCSPGSLGRCGD-IVSHSSLAUG 647
Qy      657  AVVGIILVVVLGVVFGILIKRROOKIRKYIMRBLLOETELVEPLTPSGAMPNQARILK 716
Db      648  LVSGLLITIVALLIVVLLRRRIK-RKRTIRCLLQEKELVEPLTPSGQAPNQAFRLIK 706
Qy      717  ETELKRVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKILDEAYVMAG 776
Db      707  ETEFKDVRVLSGAFGVYKGLWNPGENIRIPVAIKVLRENTSPKQNVQVLEAYVMAS 766
Qy      777  VGSPPVYSRLIGICLTSTVQLVLTQMPYGCILLDHVRENRLGSLQDILLNMCQIAKMSYL 836
Db      767  VDRPHVCRLLGICLTSAVQLVLTQMPYGCILLDVVRQHERICGQWLLNMCVQIAKGNYL 826
Qy      837  EDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPDKWNALESILR 896
Db      827  EERHLVHRDLAARNVLLKNPHVKITDFGLSKLLTADAEKEYQADGGKVPDKWNALESILQ 886
Qy      897  RRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMIW 956
Db      887  WTTHOSDWSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLPPOPICTIEVYMIIL 946
Qy      957  KCMWIDSECRPRELVSERFMRARDPQRFVWIQNEDLGPASPLDSTFYRSLLEDDDDMGD 1016
Db      947  KCMWIDPSRPRELVSERFMRARDPSRYLVIOG---NLPSLSDRFLFSRLSSDD--D 1001
Qy      1017  LVDAEYLVPOQGFCDPAPGAGGVHHRHSSTSSGSGDITGLGLEPSEERAPSRPLA 1076
Db      1002  VVDAEYLLPYKRI-----NRQGS-----EPCI 1024
Qy      1077  PSEGAGSDVFDGDLGMAKGLQSLPHDPSLQRYSEDTVP-PLPSETDGYAPLTCSP 1135
Db      1025  PPTGH-----PVRENSITURNISDPTQNALEKDLQGH----- 1056
Qy      1136  QPEYVNPQVRPOP-----PSPRE-----GPLP-AARPAGATLERAKTLSPKNGVVK 1182
Db      1057  --EYVNPQGETSRLSDIYNPNYEDLTDGMPVLSLSQEAETNFSRPEVLNTNQSL-- 1112
Qy      1183  DVFAFGAVENPEYLTQGAAPQPHPPAFSPAFNLYWDQDPPPERGAPPSTFKGTPT 1242
Db      1113  -PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTGNGMFLPA 1149
Qy      1243  AENPEYLG 1250
Db      1150  AENLEYLG 1157

ID_ERB3_HUMAN          STANDARD;          PRT;          1342 AA.
AC      P21860;
DT      01-MAY-1991 (Rel. 18, Created)
DT      01-MAY-1991 (Rel. 18, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE      (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN      ERB3 OR HER3.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90083234; PubMed=2687875;
RA      Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT      "Isolation and characterization of ERB3, a third member of the
RT      ERBB/epidermal growth factor receptor family: evidence for
RT      overexpression in a subset of human mammary tumors."
RL      Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90311312; PubMed=2164210;

```

```

RA      Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA      Todaro G.J., Shoyab M.;
RT      "Molecular cloning and expression of an additional epidermal growth
RT      factor receptor-related gene.";
RL      Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN      [3]
RP      SEQUENCE FROM N.A. (SHORT FORM).
RC      TISSUE=Placenta;
RX      MEDLINE=93282822; PubMed=7685162;
RA      Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT      "c-erbB3 gene encodes secreted as well as a transmembrane receptor
RT      tyrosine kinase.";
RL      Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC      -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAk.
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC      (POTENTIAL).
CC      -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC      SECRETED (SHORT FORM).
CC      -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC      SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC      -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC      -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC      SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC      -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC      AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC      PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC      -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
Cc      EMBL; M29366; AAA35790.1; -.
Cc      EMBL; M34309; AAA35979.1; -.
Cc      EMBL; S61953; AAB26935.1; -.
Cc      PIR; A36223; A36223.
Cc      HSPSP; P11362; LFQK.
Cc      Genew; HGNC:3431; ERBB3.
Cc      MIM; 190151; -.
Cc      InterPro; IPR000494; EGFR_L_domain.
Cc      InterPro; IPR000719; Euk_pkinase.
Cc      InterPro; IPR002174; Furin-like.
Cc      InterPro; IPR001245; Tyr_pkinase.
Cc      Pfam; PF00069; pkinase; 1.
Cc      Pfam; PF00757; Furin-like; 1.
Cc      Pfam; PF01030; Recep_L_domain; 2.
Cc      ProDom; PD000001; Euk_pkinase; 1.
Cc      SMART; SM00261; FU_3.
Cc      SMART; SM00219; TyrcK; 1.
Cc      PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
Cc      PROSITE; PS00109; PROTEIN KINASE TYR; FALSE_NEG.
Cc      PROSITE; PSS0011; PROTEIN KINASE DOM; 1.
Cc      Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
Cc      Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
Cc      Alternative splicing.
Cc      SIGNAL 1
Cc      CHAIN 20 1342
Cc      DOMAIN 20 643
Cc      TRANSMEM 644 664
Cc      DOMAIN 665 1342
Cc      DOMAIN 709 966
Cc      NP_BIND 715 723
Cc      BINDING 742 742
Cc      ACT_SITE 834 834
Cc      DISULFID 186 194
Cc      DISULFID 190 202
Cc      -----
Cc      POTENTIAL.
Cc      RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
Cc      EXTRACELLULAR (POTENTIAL).
Cc      CYTOPLASMIC (POTENTIAL).
Cc      PROTEIN KINASE.
Cc      ATP (BY SIMILARITY).
Cc      BY SIMILARITY.
Cc      BY SIMILARITY.
Cc      BY SIMILARITY.

```


RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.H., Greaves C.H., Sierke S.L., Koland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
recombinant protein.";
RN Gene 165:279-284(1995).
[2]
RP REVISIONS TO 85: 513 AND 565.
RA Hellyer N.J., Koland J.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuroligins and their putative receptors, ErbB2 and
ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
(POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
PHOSPHATIDYLINOSITOL 3-KINASE.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; U29339; AAC28498.2; -
DR EMBL; U52530; AAC53050.1; -
DR HSSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR PRINTS; PR01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 644 662 POTENTIAL.
FT DOMAIN 663 1339 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 183 259 CYS-RICH.
FT DOMAIN 707 964 PROTEIN KINASE.
FT NP_BIND 713 721 ATP (BY SIMILARITY).
FT BINDING 740 740 ATP (BY SIMILARITY).
FT ACT_SITE 832 832 BY SIMILARITY.
FT DISULFID 186 194 BY SIMILARITY.
FT DISULFID 190 202 BY SIMILARITY.
FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.

FT	DISULFID	231	243	BY SIMILARITY.
FT	DISULFID	246	255	BY SIMILARITY.
FT	DISULFID	259	286	BY SIMILARITY.
FT	DISULFID	290	301	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	323	327	BY SIMILARITY.
FT	DISULFID	500	509	BY SIMILARITY.
FT	DISULFID	504	517	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	533	549	BY SIMILARITY.
FT	DISULFID	556	573	BY SIMILARITY.
FT	DISULFID	576	585	BY SIMILARITY.
FT	DISULFID	589	610	BY SIMILARITY.
FT	DISULFID	613	621	BY SIMILARITY.
FT	DISULFID	617	629	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	408	408	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	414	414	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	469	469	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	522	522	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	566	566	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT	1028	1028	L -> P (IN REF. 3).
SQ	SEQUENCE	1339	AA; 147545	MM; 0AASF2402BBFDF1E CRC64;
Query Match 34.2%; Score 2327.5; DB 1; Length 1339;				
Best Local Similarity 40.7%; Pred. No. 9.2e-118;				
Matches 524; Conservative 169; Mismatches 429; Indels 167; Gaps 36;				
QY	3	LAALCRWGLLALLPPGAA	---STQVCTGTDMLKRLPASPETHLDMLRHLRYOGCVOVQCN	59
DB	7	LQVLC	----FLUSLARGSEMNSQAVCPGTUNGUSVTGDADNQYQTLVKLYEKCEVVMGN	62
QY	60	LELTYPNTASLSFLQDIQEQGYVLI	AHNOVRQVPLQRRLRIVRGTTQFEDNYALAVLDN	119
DB	63	LEIVLTGHNADLSFLQWIREVTGYV	LNVAMNFSVLPNLRVVRGTQVYDGFAPFM--	120
QY	120	GDPLNNTPTVGTASPGGURELQRL	SLTEILKGVLIQRNPQLCYOQTLWKDOI	179
DB	121	---LNYNT	---NSSHALRQLKFTQLTBILSGGVYIEKNKDKLCHMDTIDWRDVRVR	170
QY	180	ALTLDITNRSRACHPCSPMGKSR	CGESSDDCSLTETVCAGGC-ARCKGPLPTDCCHE	238
DB	171	GAEIVVKNNGANGCPCPEVCEK	G-RCWGGPDDCQLTKTICAPQCNGRCFGPNPQCCHD	229
QY	239	OCAAGCTGPKHSKSDCLACLFH	NSHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA	298
DB	230	ECAGCGSPQDTDCFACRRFND	SGACVPCRCPEPLVYNKLTFOLEPNPHTKYGGVCVAS	289
QY	299	CPNYLSTDVGSCTLVCPHMQE	VTABDGTORCEKSKPCARVCVGL--GMQYIKANSKP	356
DB	290	CPHNFV-VDTQFCVRACPPDK	MEVD-KHGLKMCPCGGCLCPKACEGTSGSRYQTVDSN	347
QY	357	IGITELEFAGCKKIFGSLAF	PESDGDPASNTAPLOEQLOVPETLEITCYLISAWP	416
DB	348	10----	GVNCTKILGNLDFLTGVLNVPFMHKIPALDPEKLVFRTVREITGLNIQSWP	403
QY	417	DSLPLSVFQNLQVIRGIRILH	NGAYS-LTLOGLGISWLGRLSLRELISGLSLALIHNNTHLC	475
DB	404	PHMNFVSFSLNTTIGGRSL	NYNRGFSLLIMKNLNTVSLGFRSLKSLISAGRVYISANQQLC	463
QY	476	FVHTVPDQLFRNPHQALLHT	A-NRPECEVGEGLACHQOLCARGCHWGPPTQCVNCSOF	534
DB	464	YHSLNWTLLRGPSEERLDI	KYDRPLGECLAERKVCPLCSCSGGCGWGPQGLSCRNY	523
QY	535	LRGQECVEECRVLOGLPRE	YVNAHCLCHEPCQNGSVTCFGEAOCVACAHYKQDP	594
DB	524	SREGVCVTHCNFPLOQEP	REFVHEAQCFSCHPECLPMEGTSTCNGSGSDACARCAHFRDGP	583
QY	595	FCVARCPQYIKANSKFIG	---ITELPDEGACQPCPINCTHSC--VLDLDDKGCPEAORA	648


```
Db 603 IADCGVISNAYK--FDNRCKICHPECR-----TCNGAGADHCQCVHVRDQCHQVSC 654
Qy 601 POYIKANSK-----FIGITEL----- 616
Db 655 PKN-KYNDGVCRECHATCGCTGPKDTIGGACTTCNLAIINNDATVKKRCLLKDKDKCPD 713
Qy 617 -----PDEEGACQP-----CPI-----NCTH----- 632
Db 714 GYFWEYVHPQEGSLKPLAGRAVCRKCHPLCELCNTNYGHEQVCSKCTHYKRRQECETEC 773
Qy 633 -----SC-----VDLDKG-----C 642
Db 774 PADHYTDEQRECFQHPCENGCTGPGADCKSCRNFKLFANETGYPVNSTMFNCTSKC 833
Qy 643 PAEOR-----ASPLTS-----IVSAVVGILLVVLGVVFGILI 675
Db 834 PLEMRHVNYQYTAIGPYCAASPPRSKKTANLDVNMFIITGAVLVPTICILCVV--TVI 891
Qy 676 KRQOKIRKYT--MRRLQETELVEPLTPSGAMPNOAQRILKETELRKVKVLGSGAGT 733
Db 892 CRQOKAKKETVMTMALSGCEDSEPLRPSNIGANLKLRIKIVKDAELRKGVLGAGFR 951
Qy 734 YKGIWIPDGENVKIPVAKVLRENTSPKANKEILDEAYVMAGVSPVVSRLIGTCLTST 793
Db 952 YKGVVWPEGENVKIPVAKELKSTGAESSEEFUREAYINASEEHVNLKLLAVCMSSQ 1011
Qy 794 VOLVTOLMPYGLLDHVRNRLGSDOLLNWCQIAKGMYSLEVDRLVHRDLAARNVLV 853
Db 1012 MMLITQLMPLGLLDYVNRNRDKIGSKALLNMSTQIAKGMYSLEKRLVHRDLAARNVLV 1071
Qy 854 KSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRRFTHSDVWSGVTVVW 913
Db 1072 QTPSLVKITDFGLAKLLSSDSNEYKAAGKPKIKWALECIRNVRFTSKSDVWAFGVTVI 1131
Qy 914 ELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVRCWMIDSECRPRFREL 973
Db 1132 ELLTFGQPHENIPAKDIPDLIEVLGKLEQPEICSDIYCTLLSCWHLDAAMRFTFKLT 1191
Qy 974 SEFSRMARDPQRFVVIQNEGLG--PASPLDSTFFRSLLLEDP--DMGDLVDAEYLVVQ 1028
Db 1192 TVFAEFARDGRLYLAIQDKFTRLPA-----YTSQDEKDLIRKLAPTITDGEAIAKPD 1244
Qy 1029 GFPCPDPAAGGMVHRHRSSTRSGGDLTLGLEPSEEAR-----RSPLAPSEGAG 1082
Db 1245 DYLOPKAAPGFS-----HRTDCT-----DEMPKLNRYCKDPSNKNSTG 1283
Qy 1083 SDVFDG--DLGMAAKGLQSLPHTDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQPEY 1139
Db 1284 DDERDSSAREVGVGNLR-----LDLPVDEDDYLMPTCQPGNN 1321
Qy 1140 VNQPDVVRPQPPREGPLPAARPAAGATLERAKTILSPGKNGVVKVDVAFGGAIVENPEYL-- 1197
Db 1322 NNNNN-----NPNQNNMAAVGVAAGYM-----DLIGVPVSDVNPEYLLN 1360
Qy 1198 --TPQGAAPQPH-----PPAFSP-AFDNLYWD 1224
Db 1361 AQLGVGESPIQTIGIPWVGPGTMEVKVPMVPGSEPTSSDHEYND 1408
```

RESULT 12

```
ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (BC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726 (1985).
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -|- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -|- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M10066; AAA448763.1; ALT_INIT.
CC PIR; A00643; TVCHLV.
CC PIR; B00643; TVFLV.
CC HSP; P11362; IFGK.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase.1
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE; PS00011; PROTEIN KINASE DOM; 1.
CC F. Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC DOMAIN 132 399 PROTEIN KINASE.
CC NP_BIND 138 146 ATP (BY SIMILARITY).
CC BINDING 165 165 ATP (BY SIMILARITY).
CC ACT_SITE 257 257 BY SIMILARITY.
CC SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;
CC -----
Query Match 25.3%; Score 1719.5; DB 1; Length 634;
Best Local Similarity 51.8%; Pred. No. 1.8e-85;
Matches 366; Conservative 79; Mismatches 139; Indels 123; Gaps 17;
Qy 587 CAHYKDPFCVRCPOYIKANSKFIGITELPDEBACQPCPINCTHSCVDLDDKCPAEQ 646
Db 3 CAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCOLCHPNCTRCCKGPLEGCP--- 58
Qy 647 RASPLTSIVSAVY-GILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTSGA 705
Db 59 NGSKTSIAAGVVGGLLCLVVGIGIGLYLRRR-HIVRKRTLRLLOERELVEPLTSGE 117
Qy 706 MPNQAMRILKETELRKVKVLGSGAFGTGVKGIWIPDGENVKIPVAKVLRENTSPKANK 765
Db 118 APNQAHLRLIKETEPKVKVVGSGAFGTGVKGIWIPGEKVKIPVAKLEATSPKANK 177
Qy 766 EILDEAYVMAGVSPVVSRLIGTCLTSTVOLVTLQMLPYGCLLDHVRNRLGSDQLLNW 825
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVOLVTLQMLPYGCLLDVIREHKDNIGSQYLLNW 237
Qy 826 CMQIAKGMYSLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 885
Db 238 CVQIAKGMYLEERRLVHRDLAARNVLVKTPOHKITDFGLAKLGADEKEYHAEGKVP 297
Qy 886 IKWMALESILRRRFTHSDVWSGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 945
```

```
Db      298 IKWMALESILHRIYTHQSDVMSGYTVWELMTFGSKPYDGPASEISSVLEKGERLPQP 357
Qy      946 ICTIDVYIMVKCMWIDSECRPRFRELVSFSEMRMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db      358 ICTIDVYIMVKCMWIDADSRPKFRELIAEFSKWARDPPRVLYIQGDERMHLPSPTDSKF 417
Qy      1005 YRSLEDDMDGLVDADAEVLYVPOGFFCPDPAPGAGGVVHHRSSSTRSGGDLTLGLE 1064
Db      418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy      1065 PSEEAAPSPL-----APSEGAGSDVFDGDLGMAAGKGLQSLTPHDPSPLOQRYSDPTVP 1119
Db      450 -----SRTPLLSLSLSSATSNNSATNCID-----RNGQHPVREDSFVQRYSSDPTGN 495
Qy      1120 LPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGFLPAARPAAGATLERAKTLSPGK 1177
Db      496 FLEESIDGFL-----PAPEYVNO--LMPKKPS-----TAMVQ 526
Qy      1178 NGVVKDVF-----AFGGAIVENPEYLTPOGGAAPQPHPPPAFSPAFDNLVY 1222
Db      527 NOIYNNISLTAISKLPMSDRYQNSHSTAVDNPYL-----NTNQSPLAKTVFESSPY 578
Qy      1223 WDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1254
Db      579 WIQSGNHQINLNDPDYQDQFLPNETKPNGLLKVAAPENPEYLRVAAP 625

RESULT 13
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID:79685;
RN [1]
RN [2]
RC STRAIN=H;
RA MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-78 (1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benaïssa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459 (1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A06644; TVI0H.
DR HSRP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; Atp-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132..399 PROTEIN KINASE.
FT NP_BIND 138..146 ATP (BY SIMILARITY).
FT BINDING 165..165 ATP (BY SIMILARITY).
FT ACT_SITE 257..257 BY SIMILARITY.
FT CONFLICT 29..29 R -> W (IN REF. 2).
FT CONFLICT 140..140 S -> F (IN REF. 2).
FT CONFLICT 146..146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76BBD006745D0609 CRC64;

Query Match 24.8%; Score 1688; DB 1; Length 604;
Best Local Similarity 51.7%; Pred. No. 8.6e-84;
Matches 357; Conservative 76; Mismatches 131; Indels 126; Gaps 16;

Qy 587 CAHYKDPFCVARGPQYIKANSKFITELPDEGACQCPINCTHSCVLDLDDKCGPAEQ 646
Db 3 CAHFTDGPCHVCAPAGVLGENDTL-VKRYADANAVCQLCHPNCRTGCKGPGLEGCP--- 58
Qy 647 RASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTMRLLOELTELVEPLTPSCA 705
Db 59 NGSKTPTSAAGVWGGLCLVVGVLGIGLYLRRR-HIVKRTLRLRLQRELVEPLTPSGE 117
Qy 706 MPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSFKANK 765
Db 118 APNOAHLRLKETEFKVKVLGSGAFGIYKGLWIPEGEKVIPIVAIKELRATSPKANK 177
Qy 766 EILDEAYVNAVGVSPYVSRLLIGLICLTSTVQLVTQLMYPYGCCLLDHVRNRRIGSODLLNW 825
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMYPYGCCLLDYIREHKDNIGSQVLLNW 237
Qy 826 CMOAKGMSYLEDVLRHDLAARNVLKSNHVKITDFGLARLLDIDETEYHAGGKVP 885
Db 238 CVQAKGNVLEERRLVHRDLAARNVLKTPHVKITDFGLAKLLGADEKEYHAEGKVP 297
Qy 886 IKWMALESILRRRFTHQSDVMSGYTVWELMTFGAKPYDGPAREIPOLLEKGERLPQP 945
Db 298 IKWMALESILHRIYTHQSDVMSGYTVWELMTFGSKPYDGPASEISSVLEKGERLPQP 357
Qy 946 ICTIDVYIMVKCMWIDSECRPRFRELVSFSEMRMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYIMVKCMWIDADSRPKFRELIAEFSKWARDPPRVLYIQGDERMHLPSPTDSKF 417
Qy 1005 YRSLEDDMDGLVDADAEVLYVPOGFFCPDPAPGAGGVVHHRSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1065 PSEEAAPSPL-----APSEGAGSDVFDGDLGMAAGKGLQSLTPHDPSPLOQRYSDPTVP 1119
Db 450 -----SRTPLLSLSLSSATSNNSATNCID-----RNGQHPVREDSFVQRYSSDPTGN 495
Qy 1120 LPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGFLPAARPAAGATLERAKTLSPGK 1177
Db 496 FLEESIDGFL-----PAPEYVNO--LMPKKPS----- 524
Qy 1178 NGVVKDVF-----AFGGAIVENPEYLTPOGGAAPQPHPPPAFSPAFDNLVY 1222
Db 527 NOIYNNISLTAISKLPMSDRYQNSHSTAVDNPYL-----NTNQSPLAKTVF 1218
Qy 525 --VQNYINFLISLTAISKLPMSDRYQNSHSTAVDNPYL-----NTNQSPLAKTVFE 574
Qy 1219 NLYTYWDQDQPPERGAAPPSTFKGTPTAENPEY 1248
```

```

Db 575 SSPYWTQSNHQ-----INLDNPDY 594

RESULT 14
ERBB AVIEU STANDARD; PRT; 540 AA.
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein, erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI TaxID=103898;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RL cells.";
RL Mol. Cell. Biol. 6:1751-1759(1986).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M13179; AAA2401.1; -.
CC PIR: A25231; TVFVEB.
CC HSP; P11362; IFGK.
CC InterPro: IPR000719; Euk_pkinase.
CC Pfam: PF00069; pkinase; 1.
CC ProDom: PD000001; Euk_pkinase; 1.
CC SMART: SM00219; TyrcK; 1.
CC PROSITE: PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE: PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE: PS50011; PROTEIN KINASE DOM; 1.
CC Transferrase: Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC DOMAIN 132 399 PROTEIN KINASE.
CC NP_BIND 138 146 ATP (BY SIMILARITY).
CC FT BINDING 165 165 ATP (BY SIMILARITY).
CC FT ACT_SITE 257 257 BY SIMILARITY.
CC FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
CC SEQUENCE 540 AA; 60412 MW; 5853297AA068B65D CRC64;

Query Match
Best Local Similarity 54.4%; Pred. No. 4.1e-79;
Matches 337; Conservative 69; Mismatches 123; Indels 90; Gaps 14;

QY 587 CAHYKDPVPCVACRPOYIKANSKFIGITELPDEGACQPCPINCSTHSCVDLDDKGCPSAQ 646
Db 3 CAHFDGPHCVKACPAVGLGENDTL-VKPYADANAVCQLCHPCTRGCKGPGLEGCP--- 58
QY 647 RASPLTISVSAYV-GILLVVVLGVFGILIKRQOKIRKYTWRRLLQETVELPTPSGA 705
Db 59 NGSKTPSTAAGVGVLLCVVGLGILYLRRL-HIVKRTLRLLQERLVEPLTPSGE 117
QY 706 MPNQMRILKTELKRVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANK 765
Db 118 APNOAHLKTELTKKVKVLGFGAFGVYKGLWIPGEKVTPVAIKELRATSPKANK 177
QY 766 EILDVAVYMGVGSYVSRLLGICLTSTVQLTQMPYGCCLLDVHRENRGLGSQDLN 825

```

```

Db 178 EILDVAVYMASVDNPHVCRLLGLCLTSTVQLITQMLPYGCCLLDVIREHKDNIGSQVLLNW 237
QY 826 CMQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 885
Db 238 CVQIAKGMWYLEERHWHVHRDLAARNVLVKTPOHVKITDFGLAKQLGADKEKYEAGKVP 297
QY 886 IKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPP 945
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGPASEISSVLEKGERLPQPP 357
QY 946 ICTIDVYMTWKCMWIDSECRFRRELVSFMRMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYMTWKCMWIDSECRFRRELVSFMRMARDPQRFVVIQ-NEDLGPASPLDSTF 417
QY 1005 YRSLLDEDDMGDLVDAEEVLPQGGFCFDPAPAGAGVMVHRRSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGGF-----NSPST----- 449
QY 1065 PSEEBAPRSL-----APSEGAGSDVFDGDLGMAAGKGLQSLPHTDPSPLQRYSEDP 1119
Db 450 -----SRTPLLSLSLTSATNSATNCIDRNG-----H----- 476
QY 1120 LPSETDGVVAPLTCSQPEYVNPQDVRPQPPSPREGPLPAARPAGAT-LERAKTLSPGN 1178
Db 477 -PVREDGFL-----PAPEYVQ--LMPKPESTAMVQNIYISLTAISKLPIDSRVON 527
QY 1179 GWKXDVFAFGGAVENPEYL 1197
Db 528 -----SHSTAVDNPEYL 539

RESULT 15
EGFR CHICK STANDARD; PRT; 703 AA.
ID EGFR CHICK
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER
DE (Fragment)).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha.";
RL Mol. Cell. Biol. 8:1970-1978(1988).
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

or send an email to license@isb-sib.ch).

183	Db	TVLDFASNLSSCPKCHPNCTEDHCHWGAGEQNCQTLTKVI CAQQCSGRCKRGKVPSPDCCNHQ 242
240	Qy	CAAGCTGPKHSDCLACLAHFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFEGASCVTAC 299
243	Db	CAAGCTGPRESDCCLACRFRRDATCKDTCPELVLNPTTYQMDVNPPEKYSFGATCVREC 302
300	Qy	PYNLYSTDVGSCTLVCPLHNQVTAEDGTCRCEKSKBCARCYGLGHQWYIKANSKFTGI 359
303	Db	PHNYVVDHSGSVRSNCNTDTYEV--EENGVRKCKKCDGLCSKVCNGICGIGELKGIIL--INA 360
360	Qy	TELE--FAGCKKIFGSLAFLESFQDGPASNTAPLOPEQLQVPETLEEITGYLIYISAWPDS 418
361	Db	TNIDSFNCKTNGINDVSLPVALGDAFTKTLPLDPKKLDVFRTVKEISGFLLIQAWPDN 420
419	Qy	LPDLSVFQNLQVIRGRILHNGAYSITLQGLGISWLGRSLRSLBELSGSLALIHNNHLCFVH 478
421	Db	ATDLYAFENLEIIRGTRKQHGOYSLAVVNKLKIQSLGRSLRSLKBIISGDITAI MNKNKLCYAD 480
479	Qy	TVPMQDLFRNPQHALLHTANRPEDECGVEGELACHQLCARGHCWGPGTQCVNCSQFLRQ 538
481	Db	TMNWRSLFATQSQTKIIQNRNKNKCTADRHVCDPLCSDVCGWGPFPFHCFSRFFSRQK 540
539	Qy	ECVEECRVLOGLPREYVNRHCLCHPECCQPNQ---SVTCFGBADOCVACAHYKDPDF 595
541	Db	ECVKQCNILOGEPREFERSKCLPCHSECLVQNSTAYNTTCSGPGPDHCKCAHFIDGPH 600
596	Qy	CVARCPQVIKANSFVIGITELPDEEGACQPCPINCTHSCVDLDDKGCFAEORASPLTSTIV 655
601	Db	CVKACPAGVLGENDTL--VWKYADANAVCOLCHPNCTRCCKGPGLEGCP---NGSKTPSIA 656
656	Qy	SAVV--GILLVVVLGVVFGLIKRRQQKIRKYTMRRLLQETBELVEPLTP 702
657	Db	AGVVGGLLVCVWGLGICLVLR--HIVRKETLRRLQERLVEPLTP 703

Search completed: July 22, 2003, 08:46:51
Job time : 21.2304 secs

or send an email to license@ieb-sib.ch).					
CC	-----				
DR	EMBL; M20386; AAA48760.1; .				
DR	InterPro; IPR000494; EGFR_L domain.				
DR	InterPro; IPR000719; Euk_Pkinase.				
DR	InterPro; IPR002174; Furin-like.				
DR	InterPro; IPR001245; Tyr_pkinase.				
DR	Pfam; PF00757; Furin-like; 1.				
DR	Pfam; PF01030; Recep_L_domain; 2.				
DR	SMART; SM00261; FU; 4.				
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.				
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; PARTIAL.				
DR	PROSITE; PSS0011; PROTEIN_KINASE_DOM; PARTIAL.				
KW	Transmembrane; Glycoprotein; Receptor; Signal; Transferase;				
KM	Tyrosine-protein kinase; ATP-binding; Phosphorylation.				
FT	SIGNAL 1 30				
FT	CHAIN 31 >703				
FT	DOMAIN 31 654				
FT	TRANSNEM 655 667				
FT	DOMAIN 668 >703				
FT	DISULFID 197 206				
FT	DISULFID 201 214				
FT	DISULFID 222 230				
FT	DISULFID 226 238				
FT	DISULFID 239 247				
FT	DISULFID 243 255				
FT	DISULFID 258 267				
FT	DISULFID 271 298				
FT	DISULFID 302 314				
FT	DISULFID 318 333				
FT	DISULFID 336 340				
FT	DISULFID 513 522				
FT	DISULFID 517 530				
FT	DISULFID 533 542				
FT	DISULFID 546 562				
FT	DISULFID 565 581				
FT	DISULFID 569 589				
FT	DISULFID 592 601				
FT	DISULFID 605 627				
FT	DISULFID 630 638				
FT	DISULFID 634 646				
FT	CARBOHYD 134 134				
FT	CARBOHYD 190 190				
FT	CARBOHYD 200 200				
FT	CARBOHYD 359 359				
FT	CARBOHYD 368 368				
FT	CARBOHYD 420 420				
FT	CARBOHYD 573 573				
FT	CARBOHYD 578 578				
FT	CARBOHYD 613 613				
FT	CARBOHYD 633 633				
FT	CARBOHYD 648 648				
FT	NON TER 703 703				
SO	SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;				
Query Match 23.0%; Score 1565; DB 1; Length 703;					
Best Local Similarity 44.1%; Pred. No. 4.2e-77;					
Matches 312; Conservative 111; Mismatches 255; Indels 30; Gaps 14					
QY	8 RWGLLLALLPPGAA-----STVCTGTDMKLRLPASPETHLMDLRHLRYQGCVVOGNLE	61			
DB	13 RGAAVLVLLLGVALCSAVEKKVCQGTNNKLTLQGHVEDHFTSLQRMYNNECVLSNLE	72			
QY	62 LTLPTNTASLFLODIOEVQGVYLIAHNQVRQVPQLRLIVRGTLPEDNYALAVLDNGD	121			
DB	73 ITTVEHRNDLTFLKTIQEVAGYVLIALMMVDIPLNQLIIIRGNVLYDSFALAVLSNH	132			
QY	122 PLANNTPVTGASPGGLRELQRLSLEILKGGVLIQORNPOLCYQDTILMKDFHKKNQAL	181			
DB	133 -MNKTQ-----GLRELPKMLRSLEILGGVKISNNPKLCNMDDTVLWNDIIDTSRK-PL	182			
QY	182 TLID-TNRSRACHPCSPCKSGSRGWGESSEDCQSILTRTVCAAGCA-RCKGPLPTDCCHQ	239			

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-465-479-12
Perfect score: 6804
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPSYGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6129	90.1	1259	6	O18735
2	3114	45.8	1209	11	Q9GX70
3	3085	45.3	1210	11	Q9EP98
4	2716	39.9	1165	13	Q9YH40
5	2667.5	39.2	1137	13	Q9W6F6
6	2289	33.6	1328	13	P79754
7	2007.5	29.5	1433	5	Q9BIH9
8	1871	27.5	419	4	Q9UK79
9	1719	25.6	367	11	Q8R2X1
10	1720	25.3	729	15	Q86712
11	1718	25.2	567	15	Q86714
12	1697.5	24.9	412	4	Q8WYV0
13	1653.5	24.3	962	15	Q64895
14	1645	24.2	545	15	Q85468
15	1476.5	21.7	655	11	Q9WVF5
16	1460.5	21.5	643	11	Q9ERV6

17	1274	18.7	1193	5	Q9Y1X8	Q9Y1X8 ephydatia f
18	1182.5	17.4	1368	5	Q23821	Q23821 caenorhabdi
19	1169	17.2	1717	5	Q26566	Q26566 schistosoma
20	1108	16.3	527	13	Q90836	Q90836 gallus gall
21	1001.5	14.7	478	11	Q9SE80	Q9SE80 rattus norv
22	928.5	13.6	599	13	Q9PSH2	Q9PSH2 gallus gall
23	906	13.3	165	4	Q14256	Q14256 homo sapien
24	806.5	11.9	346	13	P11776	P11776 xiphophorus
25	800	11.8	176	11	O923V5	O923V5 rattus norv
26	778	11.4	435	5	O8SZW1	O8SZW1 drosophila
27	754.5	11.1	311	13	Q99162	Q99162 xiphophorus
28	734	10.8	331	4	Q9BUD7	Q9BUD7 homo sapien
29	731.5	10.8	1362	13	Q9PVZ4	Q9PVZ4 xenopus lae
30	723	10.6	149	6	Q9BG66	Q9BG66 oryctolagus
31	721	10.6	1671	5	Q9NJV5	Q9NJV5 biophalari
32	695	10.2	1418	13	O93457	O93457 scobthalmu
33	685.5	10.1	1368	13	Q8UW85	Q8UW85 paralichthy
34	672.5	9.9	1369	13	Q8UW86	Q8UW86 paralichthy
35	664.5	9.8	1472	5	Q9U5A8	Q9U5A8 bombyx mori
36	660	9.7	1358	13	O73798	O73798 xenopus lae
37	658	9.7	1412	13	Q8UW84	Q8UW84 paralichthy
38	641.5	9.4	1418	13	Q8UW83	Q8UW83 paralichthy
39	638	9.4	1245	13	Q9YGH8	Q9YGH8 scobthalmu
40	628	9.2	1371	11	Q9QVW4	Q9QVW4 rattus sp.
41	620.5	9.1	2144	5	Q9VD94	Q9VD94 drosophila
42	600	8.8	987	11	Q91YMO	Q91YMO mus musculu
43	595	8.7	987	11	Q99MR2	Q99MR2 mus musculu
44	593	8.7	935	4	Q96L35	Q96L35 homo sapien
45	587.5	8.6	1036	4	Q07912	Q07912 homo sapien

ALIGNMENTS

RESULT 1

O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Erbb-2.
OS Canis familiaris (dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbb-2 from canine mammary gland."
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA2127.1; -
DR HSSP; P11362; IFGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP_2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match		90.1%; Score 6129; DB 6; Length 1259;
Best Local Similarity		90.0%; Pred. No. 0;
Matches 1134; Conservative		42; Mismatches 78; Indels 6; Gaps 2;
QY	1	MELAAACRGLLLALLPPGAASSTOVCTGTDMLKRLPASPTHLDMLRHLHYQGCVVQGNL 60
DB	1	MELAAACRGLLLALLPPGAAGTQVCTGTDMLKRLPASPTHLDMLRHLHYQGCVVQGNL 60
QY	61	ELTYLTNASLSFLQDIQEVQGVVLIHAHQVROVPLQRIRIVRGTLQFDNRYALAVLDNG 120
DB	61	ELTYLPANASLSFLQDIQEVQGVVLIHAHQVROVPLQRIRIVRGTLQFDNRYALAVLDNG 120
QY	121	DLPLNTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCVQDTILWKDIFPHKNQLA 180
DB	121	DPLEGGIPAGAAQGGRLRELQRLSLTEILKGGVLIQRPOLCHQDTILWKDVFHKNQLA 180
QY	181	LTLIDTNRSRACHPCSPCKGSGESSEDCSLTRTVCCAGCARCKGPLTDCCHEQC 240
DB	181	LTLIDTNRFSACPSPCKADAHWCAGSSGDCSLTRTVCCAGCARCKGPPQPTDCCHEQC 240
QY	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300
DB	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTSCP 300
QY	301	YNYLSTDVGSCTLVCPHNLQEVTAEDGTQCEKSKPCARVCYGLGMQVYIKANSKFIGIT 360
DB	301	YNYLSTDVGSCTLVCPHNLQEVTAEDGTQCEKSKPCARVCYGLGMHLREVRVTSAN 360
QY	361	ELEFACKKIFGSLAFPSFDGPASNTAPLOEOLVFETLEETIGLYISAWPDSL 420
DB	361	IOEFACKKIFGSLAFPSFDGPASNTAPLOEOLVFETLEETIGLYISAWPDSL 420
QY	421	DLVSFONLQVIRGRIHNGAYSLLTQGLGISWGLSLRELGLSLIHNHHLFCFVHTV 480
DB	421	NLSVFONLQVIRGRLVHDGAYSLLTQGLGISWGLSLRELGLSLIHNHARLCFVHTV 480
QY	481	PMQDLPROVYIKANSKFIGITELECVGEGLAHOLCARGHCWGPPTQCVNCSOFLRGQEC 540
DB	481	PMQDLFRNPHQALLHSANRPEEBECVGEGLACYP-CAHGHWCWGPPTQCVNCSOFLRGQEC 539
QY	541	VVECRVQLGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFPCVARC 600
DB	540	VVECRVQLGLPREYVKDRVCLPCHSECOPONGSVTCFGEADOCVACAHYKDPFPCVARC 599
QY	601	PSGKVPDLSPYMPKPPDEBAGCQPCPINCTHSCVDLDKGPACQASPLTSIVSAVVG 660
DB	600	PSGKVPDLSPYMPKPPDEBAGCQPCPINCTHSCADLDEKGPACQASPLTSIIAAVVG 659
QY	661	ILLVVVLGVVFGILLIKRQOKIRKYTMRLLOQTELVPELTPSGAMPNQAQMRILKETEL 720
DB	660	ILLVVVLGVVGLILIKRRQOKIRKYTMRLLOQTELVPELTPSGAMPNQAQMRILKETEL 719
QY	721	RKVVLGSGAFGVYKGIWIPDGENYKIPVAIKVLRNTSPKANKILDEAYVMAGVGP 780
DB	720	RKVVLGSGAFGVYKGIWIPDGENYKIPVAIKVLRNTSPKANKILDEAYVMAGVGP 779
QY	781	YVSRLLGICLTSTVQLVTLMPYGCILLDHHVRENRGLSGQDLNWCQIAGKMSYLEDYR 840
DB	780	YVSRLLGICLTSTVQLVTLMPYGCILLDHHVRENRGLSGQDLNWCQIAGKMSYLEDYR 839
QY	841	LVRDLAARNVLKSPNNHKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRT 900
DB	840	LVRDLAARNVLKSPNNHKITDFGLARLLDIDETEHADGGKVPKWMALLESIPRRRT 899
QY	901	HOSDVMYSYVWVWELMTFCAKPDGIPAREIPDLLEKGERLPQPICTIDVYMWKCM 960
DB	900	HOSDVMYSYVWVWELMTFCAKPDGIPAREIPDLLEKGERLPQPICTIDVYMWKCM 959
QY	961	IDSECRPRELVSFSESRMARDPQRFVJQNEDLGPASPLDSTFYRSLLDEDDMDGLVDA 1020
DB	960	IDSECRPRELVAEFSRMARDPQRFVJQNEDLGPASPLDSTFYRSLLDEDDMDGLVDA 1019
QY	1021	EYLVPOQGFCCPDPAPGAGGWHHRSSSTRSGGDLTLGLPSEEEAPRPLAPSEG 1080
Query Match		45.8%; Score 3114; DB 11; Length 1209;
Best Local Similarity		50.0%; Pred. No. 4.7e-226;
Matches 639; Conservative		165; Mismatches 358; Indels 116; Gaps 28;
QY	3	LAALCRWGLLLALLPPGA-ASTQVCTGTMDKRLPASPTHLDMLRHLHYQGCVVQGNLE 61

Db 956 DADSRPKFRELLILEFSKMDARQPORYLVIQGDERMHLPSPTDSNFYRALMDEEDMEDVDA 1015
 Qy 1021 EYLVPQOGEFFCPDPAPGAGGMVHRHRSSTSRSGGDLTLGLEPSEEEAEPRLAPSEG 1080
 Db 1016 DEYLTQOGEFF-----NSPST-----SRTPLLSLS 1041
 Qy 1081 AGSDVFDGLGMGAAGKLGSLPHTDPSPLQRYSESDPTVPLPSET--DGYVAPLTCSPQPE 1138
 Db 1042 ATSN---NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEIDNDAFL-----PVPE 1091
 Qy 1139 YVNPQDVRPOPPSPREGPLPAARPAGATLERAKTLSFGKNGVVKDVPFAGCGAVENPEYL- 1197
 Db 1092 YVNQ--SVPRKPGASQVNPVVHNOPLHP-----APGRDLHYQN--PHSNAVGNPEYLN 1140
 Qy 1198 TPOGGAAPQHPHPAPSPAFDNLVYWDQ-----DP-----PERGAPSTFKGTP 1241
 Db 1141 TAO-----PTCLSSGNSPALMIQKSHOMSLNDPDYQDFFPKETKPNIGIFAG-P 1190
 Qy 1242 TAENPEYGLDVP 1254
 Db 1191 TAENAEYLRVAPP 1203

 RESULT 4
 Q9YH40
 ID Q9YH40 PRELIMINARY; PRT; 1165 AA.
 AC Q9YH40
 DT 01-MAY-1999 (TEMBLrel. 10, Created)
 DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE Receptor tyrosine kinase proto-oncogene.
 GN XMRK.
 OS Xiphophorus xiphidium.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 ON NCBI_TaxID=8086;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RX MEDLINE=98241172; PubMed=9582016;
 RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
 RA Altschmidt J., Scharltl M.;
 RT "Activation of the Xmrk proto-oncogene of xiphophorus by
 RT overexpression and mutational alterations.";
 RL Oncogene 16:1681-1690(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RA Scharltl M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U53471; AAD10500.2; -.
 DR HSP; F11362; 1FGK.
 DR InterPro: IPR000345; CytC heme bind.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR PRINTS; PR01030; Recep_L_domain; 2.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TyrKC; 1.
 DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 2.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase

SO	SEQUENCE	1165 AA; 129614 MW; 7F7E3B8D771A74E CRC64;
Qy	Query Match	39.9%; Score 2716; DB 13; Length 1165;
Db	Best Local Similarity	45.6%; Pred. No. 5.1e-196;
	Matches	582; Conservative 162; Mismatches 384; Indels 148; Gaps 31;
Qy	1	MELAALCRWGLLALLPG-AAST-----QVCTGTDMKRLRLPASPTHLDMLRLHLYQGQCV 55
Db	4	LELLEL-----LLLLLSLIGRCSTDPDRKVCQGTSNQMTM-----LDNHYLKKMKMYSGCNV 56
Qy	56	VOGNLELYLPTNASLFLQDIOEVQGVVLIHAHNOVRQVPLQRLIRVGTOLFEVDNYALA 115
Db	57	VLENLEITYQENODLSFLOSIOEVGVVLIAMNEVSTIPLVNLRLIRGNLYEGNFLL 116
Qy	116	VLDNGDPLNTPVTGASPGGLRELQLRLTEILKGVLIQNPOLCYODTILWKDIFHK 175
Db	117	VMSNYQK-NPSSP--DVYQVGLKQLQLNLTEILSGGVKSHNPLLCNVETINMWDIVDK 173
Qy	176	NNQLALTLDINRSRACHPCSKMGKSGRCSESSEDQSLTFTVCAGGC-ARCKPLPTD 234
Db	174	TSNPTMNLIPHAERQCKQKOPGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRRCRCPKPID 233
Qy	235	CCHECAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTPESMPNPEGRYTFGAS 294
Db	234	CCNEHCAGCTGPRATDCLACRDFNDGTCKOTCPPKIYDIVSHQVVDNPNIKYTFGAA 293
Qy	295	CVTACPNYLSLTDVSGCTLVCLPHNQEVTAEDGTORCEKCKPCARCVYGLGM-----QYI 350
Db	294	CVKECPSYVYTE-GACVRSACSAGMLEVD-ENGRKSKPCDGVCPKVCDDGIGLSNTI 351
Qy	351	KANSKFIGITELEFAGCKKIGSLAFLESDDGDPASNTAPLOPQLQVFLEITGYL 410
Db	352	AVNSTNIG-----SPSNCTKINGDIILNRNSFEGDPHYKIGMPDPEHLNLTITVKEITGYL 407
Qy	411	YISAWPSLPLSLFQNLQVIRGILHNGAYS-LTQGLGISWLGSLRLSLRGSLGLAIH 469
Db	408	VIMWMPENWTSLSVFQNLLEIRGTRTFSGRSFVVVVQVSHQWLGLSLKEVSAGNVIK 467
Qy	470	HNTHLCFVHTVPMQLFRQYIKANSKFTIGITELECVGEGGLACHQLCARGHCWGPPTQCV 529
Db	468	NTPOLRVASTINWRRLFRSEDSQ-----IEYDARTENQCNBESDGCWGPPTMVCV 520
Qy	530	NCSQFLRGQECVSCRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQVCAAH 589
Db	521	SCLHVDGRGRCVASCNLLQGEPREAQVDGRVCQHQECLVQDLSLTCYGPGPANCSKAA 580
Qy	590	YKDPFPCVARGPSGVKPLSYMPIWKFPDEBACQPCPINCTHSCVDLDDKGCPEAQRAS 649
Db	581	FQDGPQCI-PRCPHGMGLGDGDL-IWKYADKMGQCPQCHQCTQCGSGPLSGCRGD-IVS 638
Qy	650	PLTSIVSVAVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQ 709
Db	639	HSSLAVALVSGLLITIVALLIVVLLRRRIK-RKRTIRRLIQEKELVEPLTPSGQAPNQ 697
Qy	710	AOMRILKETELKVKVGLSGAGFTYKIGIWDGENVKIPVAIKVLRNTSPKANKEILD 769
Db	698	AFLRLKETEFKORVLGSGAGFTYKGLMPPDGENIRIPVAIKVLRNTSPKANKEILD 757
Qy	770	EAYVMAGSPYVSRLLGICLTSTVQLVTQMLPYGCLLDHVNRNRLGSGODLLNWCQI 829
Db	758	EAYVMASVDHPVHCRLLGICLTSAVQLVTQMLPYGCLLDYVQHQERICGQWLLNWCQI 817
Qy	830	AKGMSYLEDVRLVRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGKGVPIKWM 889
Db	818	AKGMNYLEERLVRDLAARNVLLKNPNHVKITDPLGLSKLLTADEKEYQAHGKGVPIKWM 877
Qy	890	ALESILARRFTHQSDVMSYGVTVWELMTFGAKPDVGIIPAREIPOLLEGERLPOPPICTI 949
Db	878	ALESILQWYTHQSDVMSYGVTVWELMTFGSKPDVGIIPAKIASEVLENGERLPOPPICTI 937
Qy	950	DVTMIMVKCMWIDSECRPRFRELVSFRMRARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1009
Db	938	EVYMIILKCMWIDFSSRPRFRELVSFRMRARDPQRFVVIQNEIDLGPASPLDSTFYRSL 994

Qy	1010	EDDDMDLVDABEYLVYVQGGFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLSPSEBE 1069
Db	995	SSDD--DVVDADAYLL-----RYKRIN-RQGS----- 1018
Qy	1070	APRSLAPSEGAGSDVFDGLMGAAKGLQSLPHDPSPLQRYSEDPTV-PLPSETDGVV 1128
Db	1019	---EPCIPNGH-----PVRENSIALRYISDPTQNALEKDLDDGH- 1054
Qy	1129	APLTCSPQEVYVNPQDVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTLSP 1175
Db	1055	-----EYVNPQGSSTSRSLSDIYNPNVEDLTDGMPVSLSSQEAETNFSRPEYLT 1105
Qy	1176	GKNGVVDVAFAGGAVENPEYLTPOGGAPOPHPPAFSPAFDNLYYMDQDPPERGAPS 1235
Db	1106	NQSL---PLVSSGSMDDPDY---QAG-----YQAF-----LPQTGALTG 1140
Qy	1236	TFKGTPTAENPEYLG 1251
Db	1141	NGMFLPAAENLEYLGL 1156
RESULT 5		
Qy	Q9W6F6	PRELIMINARY; PRT; 1137 AA.
AC	Q9W6F6;	
DT	01-NOV-1999	(TREMBLrel. 12, Created)
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)
DE	Receptor tyrosine kinase (Fragment).	
GN	ERBB4.	
OS	Gallus gallus	(Chicken).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
OC	Gallus.	
OX	NCBI_taxID=9031;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=HINDBRAIN;	
RX	MEDLINE=99263203; PubMed=10328884;	
RA	Dixon M., Lumsden A.;	
RT	"Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in embryonic chick hindbrain."	
RL	Mol. Cell. Neurosci. 13:237-258(1999).	
DR	EMBL; AF121963; AAD31764.1; -	
DR	HSSP; P11362; IFCK.	
DR	InterPro; IPR000494; EGFR_L domain.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001368; TNFR_c6.	
DR	InterPro; IPR001245; Tyr_pkinase.	
DR	InterPro; IPR004019; YLP_motif.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep_L_domain; 1.	
DR	Pfam; PF02757; YLP; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; Tyrc; 1.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	
DR	PROSITE; PS00111; PROTEIN KINASE DOM; 1.	
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.	
DR	PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.	
KW	Kinase; Tyrosine-protein kinase.	
FT	NON_TER	
SO	SEQUENCE	1137 AA; 127927 MW; 4D616436F87DC84F CRC64;
Query Match		
Best Local Similarity		
Matches		
39.2%; Score 2667.5; DB 13; Length 1137;		
46.6%; Pred. No. 2.2e-192;		
534; Conservative 172; Mismatches 350; Indels 89; Gaps 28;		
Qy	161	LCYODTILWKDIFHKNQLALTLIDTNRSRACHPCSKMGKRCWGSESDCSLRTVC 220


```
QY 489 YIKANSKEIGITE-----LECVGLACHQLCARGHCWPGPTQCVNCSQFLRGQECVBE 543
DB 446 -IKSSDHEVMVQKRNATECHEEGECSEQSKAGCWKGPEQCLECKNWKYKCKLDS 504
QY 544 CRVLOGLPREY-VNARHCLPCHPECPONGSVTCFPEADQCACAHYKDPFPFCVARCP- 601
DB 505 CK---SLPRLYSVDSKTCODCHOECKD-----FCYGPNEEDNCGSMNVKDGRCFVABECP 556
QY 602 -----SGVKPDLSPVMPKPEPD--- 618
DB 557 TTHAMNGTCINCHKTCVGRGPRDTIADPGCISCDKAIIGSDAKTERCLMKDESPDGYV 616
QY 619 -----BEG----- 621
DB 617 SDYVLQEBGLKQLSGKAVCRKCHPRCKKCTGYGPFHQEQCBCTGYKKGEQCEDECPQDF 676
QY 622 -----ACQPCPINCT-----HSCVDL-----KGCPAQ----- 646
DB 677 YANEETRICLPCHQECRGCHGLGDDHHCNRNLKLFEGDPYDNATFTTCSVNCPSHPYKR 736
QY 647 -----RASPLTSIVSAVVGILLVVLGVVFGI---LIKRRQKI 682
DB 737 PPOEAGKIPYCSADMSQGLRIEPTQVKIVGSMVMAILLVCVFGIAFLVFSRHKKK 796
QY 683 RKYTHRRLLQETELVEPLTPSGAMPNQAMRILKETELRKVKVLSGAGFTGYKGIWIPD 742
DB 797 DAVKMTMALAGCEDSEPLRPSNVGPNLTKLRIKEAIRRGVLGAGAFGRVFKGVWMP 856
QY 743 GENKIPVIAIKVLENTSPKANKEILDENYVWAGVSPVSRLLGICLTSTVOLTQLMP 802
DB 857 GESVKIPVIAIKVLENTSPKANKEILDENYVWAGVSPVSRLLGICLTSTVOLTQLMP 916
QY 803 YGCLLDHVRNCRGLSGODLNCMCQIAKMSYLEDRVLVHRDLAARNVLKSPNHNKIT 862
DB 917 LGCLLDYRNKDKTIGSKALLNWSQIARGMAYLEERLVRDLAARNVLVQTPSCVKIT 976
QY 863 DFLARLLDIDETEHADGKVPKIKWMALESILRRRFTHQSDVWSYGYVTVMELMTFGAKP 922
DB 977 VFLAKLLDFDSDEYRAAGGKMPIKWALECIRHVRFTSKSDVWAFGTTIMELLYGARP 1036
QY 923 YDGPAREIPDLLEGERLPQPICTIDVYMIWVCMWIDSECRFRRELVSERFARMARD 982
DB 1037 YENVPKDVPELIEIGHKLQPDICSLDYVCISSWLDADARPTFKQLAETFAEKARD 1096
QY 983 PORFVVIQNEDLGPASPLDSTFYRSLLDEDDMDLV----- 1018
DB 1097 PORYLMI-----PGDKFMRPLPSYTNQDEKDLIRTLAPVMAAAAAAAGASNVUV 1147
QY 1019 -----DABEYLVOQGFCCPDPAQAGGMVHRRHSSSTRSGGDLTLGLEPSEEAPRS 1073
DB 1148 PSTIAETDEYLQPKTRPSIMLPGPSA-----VEPS-DEMPKS 1183
QY 1074 -----PLAP-----SEGASDVFDDGLGMAKGLQSLTHDPSPLORYSEDPVPLPSET 1124
DB 1184 LRYCKDPLKPDDETHGKEV-----GVGGIR-----LNLPLDE 1217
QY 1125 DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAAPAGATLERAKTSLSPGKNWGVKDV 1184
DB 1218 DDYLMF-TCQSQ-----NQS-----TFG-----YMDL 1238
QY 1185 FAFGAVENPEYL-----TPGGAAPOHPHPAPFSPAPDNLYYWDQDPPPERGAPPS 1235
DB 1239 IGVPAVDNPEYLMGSTQAIAGLAQSGMG---PHTPP-----PPN 1275
QY 1236 TFKGTPTAENPE 1247
DB 1276 TPNGMPTHQHSQ 1287
```

RESULT 8

```
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
```

```
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF177761; AAS56009.2; -
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like; 1.
DR SMART: SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D0030C CRC64;

Query Match 27.5%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 7.2e-133;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASQVCTGCTDMKRLRPASPTHLDMLRHLVQGCQVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASQVCTGCTDMKRLRPASPTHLDMLRHLVQGCQVQGNL 60
QY 61 ELTYLPTNASLFLQDIOEVQGYVLIHNOVRQVPLQRLRIVRGTLQFEDNYVALVDNG 120
DB 61 ELTYLPTNASLFLQDIOEVQGYVLIHNOVRQVPLQRLRIVRGTLQFEDNYVALVDNG 120
QY 121 DPLNNTTPTVGTASPGGLRELQRLSITELKGGVLQRPOLCYQDTILWKDIFHKNNQLA 180
DB 121 DPLNNTTPTVGTASPGGLRELQRLSITELKGGVLQRPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPCKGSRGSESDCSLRTVVCAGGCARCKGLPTDCCHQC 240
DB 181 LTLIDTNRSRACHPCSPCKGSRGSESDCSLRTVVCAGGCARCKGLPTDCCHQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNQVETADGTQRCCKSKPCARVCYGL 345
DB 301 YNYLSTDVGSCTLVCPHLNQVETADGTQRCCKSKPCARVCYGL 345

RESULT 9
QY Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
```



```
DR EMBL; BC027080.1; --
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match      25.6%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 5.5e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 889 MALESILRRRFTTHOSDWSYGVTVWELMTFGAKYDGIPIAREIPDLLEKGERLPPICT 948
Db 1 MALESILRRRFTTHOSDWSYGVTVWELMTFGAKYDGIPIAREIPDLLEKGERLPPICT 60

QY 949 IDVTMIMVKWIMIDSECRPRELVSFSESRNARDPQRFVQIQLNEDLGPASLDSTFYRSL 1008
Db 61 IDVTMIMVKWIMIDSECRPRELVSFSESRNARDPQRFVQIQLNEDLGPASLDSTFYRSL 120

QY 1009 LEDDMDGLVDAEYLVPQGFPCDPAPGAGGMVHRHRSRSTSRSGGDLTLGLEPSEE 1068
Db 121 LEDDMDGLVDAEYLVPQGFPCDPALGTGSTAHRHRSRSTSRSGGDLTLGLEPSEE 180

QY 1069 EAPRSPLAPSEGAGSDVFDGLGMAAKGLOSLTHDPSPQLQRYSEDPTVPLPSETDGYV 1128
Db 181 EAPRSPLAPSEGAGSDVFDGLAVGVTGKLSLPHOLSPQLQRYSEDPTVPLPSETDGYV 240

QY 1129 APLTCSQPEVYNQDVRPQPPSPREGPLPAAPAGATLERAKTLPCKNGVWKFVAFG 1188
Db 241 APLTCSQPEVYNQDVRPQPPSPREGPLPAAPAGATLERAKTLPCKNGVWKFVAFG 300

QY 1189 GAVENPEYLPQGGAAQPPHPPAFSPAFNLYYWDQDPPRGERGAPPSTFKGTPTAENPEY 1248
Db 301 GAVENPEYLPQGGAAQPPHPPAFSPAFNLYYWDQDPPRGERGAPPSTFKGTPTAENPEY 360

QY 1249 LGLDVPV 1255
Db 361 LGLDVPV 367

RESULT 10
Q86712 ID Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviridae; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; --
DR HSP; P03322; IAGS.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFE1D63 CRC64;
```

```
Query Match      25.3%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 4.2e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

QY 569 PQNGSVTCFGEADOCVACAHYKDPFPCVACRCPGVKPDLSYMPILWKFPEDEGACQCPPI 628
Db 141 PEETATPKTGP--DHCWKCAHFDGPHCVKACPAAGVIGENDTL-VWKYADANAVCOLCHP 197

QY 629 NCHTSCVDLDDKGPACQASPLTSIVSAVV-GILLVVVLGVFGILIKRRQOKIRKTYM 687
Db 198 NCTRGCKGPGLEGCP--NGSKTPIAAGVVVGLLCLVVGGLGILGLYLR--HIVRKRLL 253

QY 688 RLLQETELVEPLTPSCAMPNQAMRILKETELRKVKVLGSGAGFTVYKGIWIDGENVK 747
Db 254 RLLQERELVEPLTPSGEAPNQAHRLIKETEFKVKVVLGSGAGFTVYKGLWIPEGEVK 313

QY 748 IPVAIKVLRNTSPKANKEILDEAYVWAGVSPVSRLLGICLTSTVOLVTOLMPYGCCL 807
Db 314 IPVAIKELREATSPKANKEILDEAYVWASVDNPRVCRLLGICLTSTVOLITOLMPYGCCL 373

QY 808 DHVRENRLGSGDOLLNWCMIKAGMSYLEDRVLVHRDLAARNVLKSPNHVKITDFGLA 867
Db 374 DYIREHKDNIGSOYLLNWCMIKAGMYLEERRLVHRDLAARNVLKTPQHVKITDFGLA 433

QY 868 RLDDIDETHADGKGVPIKMALESILRRRFTHOSDWSYGVTVWELMTGAKPYDGIP 927
Db 434 KLLGADEKEYHAEKGKVPKMALESILHRIYTHOSDWSYGVTVWELMTGSKPYDGIP 493

QY 928 AREIPDLEKGERLPPICTIDVYMTIMVKWIMIDSECRPRELVSFSESRNARDPQRFV 987
Db 494 ASELSVLEKGERLPPICTIDVYMTIMVKWIMIDSECRPRELVSFSESRNARDPQRFV 553

QY 988 VIQ-NEDLGPASPLDSTFYRSLLEDDMDGLVDAEYLVPQGFPCDPAPGAGMVHHR 1046
Db 554 VIQDERMHLPSPTDSKFYRTLMEEDMEDIVDAEYLVPQGF----- 598

QY 1047 HRSSTRSGGDLTLGLEPSEEEAPRSP-----APSEGAGSVDFDGLGMAAKGLOSL 1101
Db 599 -NSPST-----SRTPLSSLSATSNNSATNCID-----RNGQGH 631

QY 1102 PTHDPSPLQRYSEDPTVPLPSET--DCGVAPLTCSQPEVYNQDVRPQPPSPREGPLPA 1159
Db 632 PVREDTSFQRYSSDPTGNFLEESIDGFL-----PAPEYVQ--LMPKFPS----- 675

QY 1160 ARPAGATLERAKTLPCKNGVWKFV-----AFGGAVENPEYL 1197
Db 676 -----TAMVQNQIYNNISLTAKLPMDRSYQNSHSTAVDNPEYL 715

RESULT 11
Q86714 ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviridae; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; --
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
```

```
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON TER
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 25.2%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 4.1e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 578 GPEADQCACAHYDPFCVACRSGVKPDLISYPIWKFPEDEACQPCINCHTSCVDL 637
Db 1 GP--DHCKMCAHFIDGHCVCACAGVLGENDTL-VMKYADANAVCQLCHPNCRTGCKGP 57
QY 638 DDKGCPAEQASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYMRRLQETEL 696
Db 58 GLEGCP---NGSKTPSTAAGVVGULLCLVVGLGIGLYLRR-HIVKRTLRLLQEREL 113
QY 697 VEPLTPSGAMPNQOMRILKETELRKVKVLGSGAGFTYKGIWIPDGENKIPVAIKVLR 756
Db 114 VEPLTPSGEAPNQAHRIKETEPKVKVLGSGAGFTYKGLWIPGEKVKIPVAIKELR 173
QY 757 ENTSPKANKEILDEAYVMAGVSGPYVSRLLGICLTSTVOLVQLMPCGCLLDHVRNRR 816
Db 174 EATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVOLVQLMPCGCLLDYIREHKN 233
QY 817 LGSODLLNWCQIAKGSYLEEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
Db 234 IGSYLLNWCQIAKGNYLEERLVHRDLAARNVLKTPQHVKITDFGLAKLGADEKE 293
QY 877 YHADGKVPKIMMALESILRRFTHQSDVMSYGYVTVMELMTFGAKPYDGIPIAREIPDLE 936
Db 294 YHAEKGKVPKIMMALESILRHRYTHQSDVMSYGYVTVMELMTFGSKPYDGIPIASEISSVLE 353
QY 937 KGERLPPOPICTIDVYIMVKCWMIDSECRPFRELVSERFMRARDPQRFVVIQ-NEDLG 995
Db 354 KGERLPPOPICTIDVYIMVKCWMIDADSRPKFELIAEFKMRARDPPRYLVVIQGDERMH 413
QY 996 PASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRRHSSSTRSG 1055
Db 414 LPSTDSKFYTLNEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
QY 1056 GGDUTLGLPSEEEAPRSPL-----APSEAGSDVFDGDLGMAAKGLQLPTHDPSPLQ 1110
Db 455 -----SRTPLLSLSATSNNNSATNCID-----RNGQHPVREDSFVQ 491
QY 1111 RYSEDPTVPLPSET--DGYVAPLTCSPQEVYVQPDVVRPPSPREGPLPAARPAATLE 1168
Db 492 RYSDPTGNFLEESIDGFL-----PAPEYVNO--LMPKKPS----- 526
QY 1169 RAKTLPSPKGVKVDVF-----AFGGAVENPEYL 1197
Db 527 ----TAMVQIQIYNNISITAKSLPMDSRVQNSHSTAVDNPYIL 566

RESULT 12
Q8WYV0
ID Q8WYV0 PRELIMINARY; PRT; 412 AA.
AC Q8WYV0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN p31659.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RT
```

```
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF18349; AAL55856.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR004019; YLP motif.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.9%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 9e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 889 MALESILRRRFTHQSDVMSYGYVTVMELMTFGAKPYDGIPIAREIPDLLEKGERLPPICT 948
Db 1 MALESILRRRFTHQSDVMSYGYVTVMELMTFGAKPYDGIPIAREIPDLLEKGERLPPICT 60
QY 949 IDVYIMVKCWMIDSECRPFRELVSERFMRARDPQRFVVIQNEDELGPASLDSTFYRSL 1008
Db 61 IDVYIMVKCWMIDSECRPFRELVSERFMRARDPQRFVVIQNEDELGPASLDSTFYRSL 120
QY 1009 LEDDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRRHSSSTRSGGDLTLGLEPSEE 1068
Db 121 LEDDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRRHSSSTRSGGDLTLGLEPSEE 180
QY 1069 EAPRSLPASEGAGSDVFDGDLGMAAKGLQLPTHDPSPLQRYSEDPVPLPSETDGYV 1128
Db 181 EAPRSLPASEGAGSDVFDGDLGMAAKGLQLPTHDPSPLQRYSEDPVPLPSETDGYV 240
QY 1129 APLTCSPPQEVYVQPDVVRPPSPREGPLPAARPAATLEBRKTLSPKGVKVDVPAFG 1188
Db 241 APLTCSPPQEVYVQPDVVRPPSPREGPLPAARPAATLEBRKTLSPKGVKVDVPAFG 300
QY 1189 GAVENPEYLTPOGGAAPQP-----HPPPA---FSPAEDNL 1220
Db 301 GAVENPEYLTPOGGAALSPDLLPSAQPSTTSITGTRTHOSGGLHPAPSKGHLRQRTST 360
QY 1221 YYWD-QDPPER-----GAPPSTFKCTPTAEN 1245
Db 361 VVWTCOCEPEGVRRSPDVSSGREGLTSAKIKRWEGPPTTSRGTCARN 410

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
DE GAG, v-erb-A, v-erb-B.
GN Avian erythroblastosis virus.
OS Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OC NCBI_TaxID=11861;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Bruskina A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B.";
```



```
Qy 996 PASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCCPDAPGAGGMVHHRSSSTRSG 1055
Db 414 LPSFTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
Qy 1056 GGDLTGLFSESEAPRSL-----APSEGAGSVDFGCDLGMGAAGLQSLPHTDPSPLQ 1110
Db 455 -----SPTLSSLSATSNNSATNCIDRNG-----H----- 481
Qy 1111 RYSEDPTVPLPSETDGVVAPLTCSPQEPYVQNPQVRPQPPSPREGPLPAARPAAGAT-LE 1169
Db 482 -----PVREDGFL-----PAPEYVNO--LMPKKSTAMVQIQIYNYISLTAKS 523
Qy 1170 AKTLPSPKGVVQKDVFAFGAVENPEYL 1197
Db 524 LPMDSRYN-----SHSTAVDNPEYL 544

RESULT 15
Q9WVF5 Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor
DE isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
(1)
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Mahle N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT Carboxy-Terminal Truncated Receptors."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
(2)
RN SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mahle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
(3)
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuoka H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Sakurai L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bareh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohseki S.,
RA Hayashizaki Y.,
```

```
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -.
DR EMBL; AK004883; BAB23641.1; -.
DR EMBL; AK004911; BAB23662.1; -.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC28CB CRC64;
```

```
Query Match 21.7%; Score 1476.5; DB 11; Length 655;
Best Local Similarity 43.9%; Pred. No. 8.8e-103;
Matches 281; Conservative 97; Mismatches 237; Indels 25; Gaps 9;
```

```
Qy 11 LLLALLPPGAA--STOVCTGTDMKRLRLPASPTHDLMLRHLHYQCQVVOGNLELYLPTN 68
Db 14 LLTALCAAGGALBEKKVCQGTSNRLTQLGTFFEDHFLSLQRMYNVCEVWLGNLLEITYVQRN 73
Qy 69 ASLSFLQDIQEVQGYVLIHQNVRQVPLQRLRIVRGTLQFEDNVVALAVLDNGDPLNNTTP 128
Db 74 YDLSFLTKITQEVAGYVLIHNTVERIPLENLQIRGNALYENTYALAILSN----- 124
Qy 129 VTGASPGGLRELQRLSLTEILKGVLIQRLPOLCYODTILWKDI---FHNKQLALTLI 184
Db 125 -YGTNRGTGLRELPMRLNQLLEILGAVRFSNNPILCNMDTIQWRDILVQNVFSNMSMDL--- 180
Qy 185 DTRNSRACHPCSPMKSGRCWGESSEDCOSLRTVTCAGGCA-RCKGPLPTDCCHCEQCAAG 243
Db 181 -QSHPSRCPKCDSPCNWSCWGGEENCQKLTIIICAAQCSHRRCGRSPDCCHNQCAAG 239
Qy 244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACPNY 303
Db 240 CTGPRESDECLVCOKFODEATCKDTCPLMLYNFTTYQMDVNPBGKVSFGATCVKCKPRNY 299
Qy 304 LSTDVSGSCTLVCPHLNQEVTAEDGTQRCBCKSKPCARVCYGLQMYIKANSKFITGITELE 363
Db 300 VTDHGSCVACGPDYEV-EEDGIRKCKKCDGCRKVCNGIGIGBEFK-DTSLINATNIK 357
Qy 364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEITGYLISAMPDLSPLD 422
Db 358 HFKYCTAISGDLHLPLVAFKGSFTRTPPLDPRELILKTVKEITGFLLIQAMPDNTDL 417
Qy 423 SVFQNLQVIRGRILHNGAYSLTLOGLGISWGLRSRLBELGSLALIHNTHLCFVHTVPW 482
Db 418 HAFENLEIRGRTKHQGFSLAVVGLNITSLGRLSLKEISDGDVVIISGNRLCYANTINW 477
Qy 483 DQLFROYIKANSKFITGITELECVGEGLAGLCHQLCARGHCWGPPTQCVNCSQFRLRGSCVE 542
Db 478 KKLFGTPTNOKTKIMNRAEKDKAVNVHVCNPLCSSGCGWGPEDPCVSCQNVSRGRCVE 537
Qy 543 ECRVLOGLPREYVNAHRLCPHCEQPNQSGVTCFGEADQCVCACAHYKDPKPPFCVARCP 602
Db 538 KCVILEGEPREFVENSECICQHCPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTC 597
Qy 603 GVKPDLSTYMPIWKFPEDEACQCPINCTHSCVDLDDKGC 642
Db 598 GIMGENTL-VWKYADANNVCHLCHANCTYGCAGPGLQGC 636
```

Search completed: July 22, 2003, 09:01:22
Job time : 53.3575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-465-479-12

Perfect score: 6804

Sequence: 1 MELAALCRWGLLLALLPPCA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6637	97.5	1255	AA192620	Human heregulin 2
2	6637	97.5	1255	AAE12130	Human tyrosine kin
3	6637	97.5	1255	AAE60167	HER2 transgene pla
4	6637	97.5	1255	AAU74545	Human HER2 (erbB2)
5	6631	97.5	1255	AAW01111	HER-2/neu protein
6	6631	97.5	1255	AAW92406	Human HER-2/neu on
7	6631	97.5	1255	AAW92406	Human HER-2/neu pr
8	6631	97.5	1255	AAW84780	Amino acid sequenc
9	6631	97.5	1255	AAW84548	Human HER-2/neu pr
10	6631	97.5	1255	AAW88267	HER2/neu amino aci

11	6631	97.5	1255	23	AAE24067	Human Her-2 protei
12	6631	97.5	1255	23	AAE20479	Human Her-2/neu pr
13	6631	97.5	1255	23	AAW51143	Human Her-2/neu on
14	6631	97.5	1255	23	AAU77114	Human Her-2/neu po
15	6588	96.8	1433	14	AAK39568	Sequence of c-erbB
16	6467	95.0	1223	23	AAU98923	Human breast cance
17	6314	92.8	1200	21	AAU821208	Human HER-2/neu pr
18	5854.5	86.0	1256	21	AAE211199	Rat HER-2/neu prot
19	5854.5	86.0	1256	23	AAW51144	Rat HER-2/neu onco
20	5827.5	85.6	1256	21	AAE21206	Mouse Her-2/neu pr
21	5827.5	85.6	1256	22	AAE21206	Amino acid sequenc
22	5827.5	85.6	1256	23	AAW51151	Mouse Her-2/neu on
23	4717	69.3	919	21	AAE21203	Human HER-2/neu fu
24	4717	69.3	919	23	AAW51148	Her-2/neu extracel
25	3977.5	58.5	920	23	AAW51152	Mouse Her-2/neu ex
26	3977.5	58.5	926	23	AAW51153	Mouse Her-2/neu ex
27	3601	52.9	712	23	AAE21204	Human HER-2/neu fu
28	3601	52.9	712	23	AAW51149	Her-2/neu extracel
29	3455	50.8	782	18	AAW19764	Her2-GM-CSF immuno
30	3453	50.7	653	21	AAE21200	Extracellular HER-
31	3453	50.7	653	23	AAW51145	Human Her-2/neu on
32	3415	50.2	645	22	AAE60408	Human ErbB2 oncopr
33	3415	50.2	645	22	AAE61593	Human ErbB2 extrac
34	3350	49.2	951	21	AAV44993	DC86FV-erbB2EC fu
35	3247	47.7	624	11	AAE08222	Extracellular port
36	3121	45.9	1210	21	AAE19259	Amino acid sequenc
37	3121	45.9	1210	21	AAV50616	Human EGF receptor
38	3121	45.9	1210	23	AAE23019	Human Her-1 protei
39	3121	45.9	1210	23	AAW50768	Human epidermal gr
40	3119	45.8	1210	22	AAE68420	Amino acid sequenc
41	3084	45.3	583	23	AAE20483	Human protein for
42	3084	45.3	587	23	AAE20481	Human protein for
43	3083	45.3	589	23	AAE20484	Human protein for
44	3083	45.3	600	23	AAE20482	Human protein for
45	3080	45.3	1210	23	ABP51768	Human epidermal gr

ALIGNMENTS

RESULT 1
A_Y92620
ID AA192620 standard; Protein; 1255 AA.

XX AC AA192620;

XX DT 10-AUG-2000 (first entry)

XX DE Human heregulin 2 (Her2).

XX KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
XX KW self-protein; cancer; breast cancer; prostate cancer;
XX KW cell-associated peptide antigen; foreign epitope.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Domain 1..173

XX FT /label= N-terminal

XX FT /note= "mature polypeptide"

XX FT Region 5..25

XX FT /label= insertion region

XX FT /note= "suitable for foreign epitope insertion"

XX FT Region 59..73

XX FT /label= insertion region

XX FT /note= "suitable for foreign epitope insertion"

XX FT Region 103..117

XX FT /label= insertion region

XX FT /note= "suitable for foreign epitope insertion"

XX FT Region 149..163

XX FT /label= insertion region

XX FT /note= "suitable for foreign epitope insertion"

XX FT Domain 174..323

QY 781 YVSRLLGICLTSTVOLTPMYPGCLLDHVRNRCRGLSGODLLNWKMOIAKMSYLEDVR 840
Db |||||
QY 781 YVSRLLGICLTSTVOLTPMYPGCLLDHVRNRCRGLSGODLLNWKMOIAKMSYLEDVR 840
Db |||||
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900
Db |||||
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900
Db |||||
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKCM 960
Db |||||
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKCM 960
Db |||||
QY 961 IDSECRPRFELVSEFSRWARDPQRFVVIQNEIDLGASPLDSTFYRSILDDDDMDGLVDA 1020
Db |||||
QY 961 IDSECRPRFELVSEFSRWARDPQRFVVIQNEIDLGASPLDSTFYRSILDDDDMDGLVDA 1020
Db |||||
QY 1021 EYLVPQOGFFCPDPAPGAGGVHRRHSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db |||||
QY 1021 EYLVPQOGFFCPDPAPGAGGVHRRHSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db |||||
QY 1081 AGSDVFDGDLGMAAGKLOSLTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db |||||
QY 1081 AGSDVFDGDLGMAAGKLOSLTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db |||||
QY 1141 NQPDVRPQPPSPREGPLPAAPAGATLERAKTLPKNGVVKDVFAGGAVENPEYLTPO 1200
Db |||||
QY 1141 NQPDVRPQPPSPREGPLPAAPAGATLERAKTLPKNGVVKDVFAGGAVENPEYLTPO 1200
Db |||||
QY 1201 GGAAPQHPHPPAFSPAFDNLYWDDPPERGAPSTFKGTPTAENPEYLGLDVPV 1255
Db |||||
QY 1201 GGAAPQHPHPPAFSPAFDNLYWDDPPERGAPSTFKGTPTAENPEYLGLDVPV 1255
Db |||||
RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 774..782
FT /note= "Antigenic epitope"
XX
XX WO200168677-A2.
XX
XX 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US40328.
XX
XX 16-MAR-2000; 2000US-0527487.
XX
XX (GENZ) GENZYME CORP.
XX
XX Nicolette CA;
XX
XX WPI; 2001-616284/71.
XX
XX N-PSDB; AAD19731.
XX
XX Novel synthetic therapeutic compound for inducing immune response and
XX for use in adoptive immunotherapy, has enhanced binding to major
XX histocompatibility molecules and enhanced immunoregulatory properties
XX

XX
PS Claim 4; Page 63-67; 69pp; English.
XX
CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;
Query Match 97.5%; Score 6637; DB 22; Length 1255;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1227; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPFGAASTQVCTGCTDMKLRLPASPETHLMLRLHYQSCVVQGNL 60
Db |||||
QY 1 MELAALCRWGLLLALLPFGAASTQVCTGCTDMKLRLPASPETHLMLRLHYQSCVVQGNL 60
Db |||||
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNQVRQVPLQRLIRVGTQLFEDNVALVDNG 120
Db |||||
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNQVRQVPLQRLIRVGTQLFEDNVALVDNG 120
Db |||||
QY 121 DPLNNTPTVTGASPGGLRELQRLSITELKGGVLTORNPOLCYQDTILWKDIFHKNQOLA 180
Db |||||
QY 121 DPLNNTPTVTGASPGGLRELQRLSITELKGGVLTORNPOLCYQDTILWKDIFHKNQOLA 180
Db |||||
QY 181 LTLIDTNRSRACHPCSPMCKGRCWGESSEDCQSLTRTVACGACARCKGPLPTDCHEQC 240
Db |||||
QY 181 LTLIDTNRSRACHPCSPMCKGRCWGESSEDCQSLTRTVACGACARCKGPLPTDCHEQC 240
Db |||||
QY 241 AAGCTGPKHSCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGYTFGASCVTACP 300
Db |||||
QY 241 AAGCTGPKHSCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGYTFGASCVTACP 300
Db |||||
QY 301 YNYLSTDVGSCCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIT 360
Db |||||
QY 301 YNYLSTDVGSCCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
Db |||||
QY 361 ELEFAGCKKIFGSLAFAPESFDGDPASNTAPLQPEQLQVFTLEBITGYLISANPDSL 420
Db |||||
QY 361 IQEFAGCKKIFGSLAFAPESFDGDPASNTAPLQPEQLQVFTLEBITGYLISANPDSL 420
Db |||||
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLISWGLRSRLRELSGLALIHNNTHLCFVHTV 480
Db |||||
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLISWGLRSRLRELSGLALIHNNTHLCFVHTV 480
Db |||||
QY 481 PWDQLFRQYIKANSKFITGITELECVGEGGLACHOLCARGHCGPQTQCVNCSQFLRGQEC 540
Db |||||
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCGPQTQCVNCSQFLRGQEC 540
Db |||||
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEVCAR 600
Db |||||
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEVCAR 600
Db |||||
QY 601 PSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTSIVSV 660
Db |||||
QY 601 PSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTSIVSV 660
Db |||||
QY 661 ILLVVLGVVFGILIKRROQKIRKYTWRRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
Db |||||
QY 661 ILLVVLGVVFGILIKRROQKIRKYTWRRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
Db |||||

Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMKMALESILRRRT 900
 Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWCWM 960
 Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWCWM 960
 Qy 961 IDSECRPRFRELVEFSRMDRDPORFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1020
 Db 961 IDSECRPRFRELVEFSRMDRDPORFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1020
 Qy 1021 EYLVPQOQFFCPDPAPGAGMVHRRSSSTRSGGDLTLGLPESEEEAPRSLAPSEG 1080
 Db 1021 EYLVPQOQFFCPDPAPGAGMVHRRSSSTRSGGDLTLGLPESEEEAPRSLAPSEG 1080
 Qy 1081 AGSDVFDDGLMGAAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDDGLMGAAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
 Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
 Qy 1201 GGAAPQPHPPAFSPAFDNLVYMDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
 Db 1201 GGAAPQPHPPAFSPAFDNLVYMDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 4

AAU74545

ID AAU74545 standard; Protein; 1255 AA.

XX AAU74545;

XX 23-APR-2002 (first entry)

XX Human HER2 (ErbB2) polypeptide.

KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glial disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macrophagal disorder; epithelial disorder;
 KW atromal disorder; blastocoealic disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.

XX Homo sapiens.

XX US2002001587-A1.

XX 03-JAN-2002.

XX 16-MAR-2001; 2001US-0811123.

XX 16-MAR-2000; 2000US-189844P.

XX 05-OCT-2000; 2000US-238327P.

XX (ERIC/) ERICKSON S.

XX (SCHW/) SCHWALL R.

XX (SLIW/) SLIWKOWSKI M.

XX Erickson S, Schwall R, Sliwkowski M;

XX WPI; 2002-163686/21.

XX N-PSDB; ABK14058.

XX Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor, ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal -

XX Example 3; Fig 7; 93pp; English.

XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.

XX SQ Sequence 1255 AA;

Query Match 97.5%; Score 6637; DB 23; Length 1255;

Best Local Similarity 97.8%; Pred. No. 0;

Matches 1227; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPFGAASSTQVCTGTDMKRLPASPEHLDMLRHLYOCQVVOGNL 60
 Db 1 MELAALCRWGLLLALLPFGAASSTQVCTGTDMKRLPASPEHLDMLRHLYOCQVVOGNL 60
 Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLI AHNQVRQVPLQRLRIVRGCTQLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIQEVQGYVLI AHNQVRQVPLQRLRIVRGCTQLFEDNYALAVLDNG 120
 Qy 121 DPLNNTPTVTGASPGGLRELOLSLTELKGGVLIQRNPOLCYQDTTLWKDIFHKNNOLA 180
 Db 121 DPLNNTPTVTGASPGGLRELOLSLTELKGGVLIQRNPOLCYQDTTLWKDIFHKNNOLA 180
 Qy 181 LTLIDNRSRACHPCSPMKCGSRGWGESSDQSLTRTVACAGGCARCKGPLTCCHEQC 240
 Db 181 LTLIDNRSRACHPCSPMKCGSRGWGESSDQSLTRTVACAGGCARCKGPLTCCHEQC 240
 Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVSGCTLVCPLHNOEVTAEQGTORCEKSKPCARVCYGLGHQYIKANSKFQIT 360
 Db 301 YNYLSTDVSGCTLVCPLHNOEVTAEQGTORCEKSKPCARVCYGLGHQYIKANSKFQIT 360
 Qy 361 ELEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLOVFETLEEITGYLYISAWPDSL 420
 Db 361 IQEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLOVFETLEEITGYLYISAWPDSL 420
 Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALIHNNTHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALIHNNTHLCFVHTV 480
 Qy 481 PWDOLFROYIKANSKFQITGITECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRQEC 540
 Db 481 PWDOLFQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALIHNNTHLCFVHTV 480
 Qy 541 VEBCRVQLGLPREYVVARHCLPCHPECPONGSVTCFGEADOCVCAHYKDPFCVARC 600
 Db 541 VEBCRVQLGLPREYVVARHCLPCHPECPONGSVTCFGEADOCVCAHYKDPFCVARC 600
 Qy 601 PGVVKPDL SYMPIWKFPDDEGACQPCP INCTHSCVDLDDKGCPEAQSRASPLTSIVSAVVG 660
 Db 601 PGVVKPDL SYMPIWKFPDDEGACQPCP INCTHSCVDLDDKGCPEAQSRASPLTSIVSAVVG 660
 Qy 661 ILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTS PKANKETLDEAYVMAGVGP 780
 Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTS PKANKETLDEAYVMAGVGP 780
 Qy 781 YVSRLLGICLTSTVQLVTQTLMPYGCILLDHVRENRRGLSGQDLNLNWCQIAKSGMYSLEDDVR 840

Db 781 YVSRLLGICLTSTVQLVTPMPYGCGLDHRVNRGRGLGSQDLNWCMIAGMSYLEVDV 840
Qy 841 LVHRDLAARNVLVSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRPT 900
Db 841 LVHRDLAARNVLVSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRPT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYIMVVKWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYIMVVKWM 960
Qy 961 IDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGASPLDSTFYRSLLDDDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGASPLDSTFYRSLLDDDDMDGLVDA 1020
Qy 1021 EYILVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Db 1021 EYILVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGKGLSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGKGLSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVDVPAFGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVDVPAFGAVENPEYLTPO 1200
Qy 1201 GGAAPQPHPPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQPHPPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 5
ID AAW01111 standard; Protein; 1255 AA.
XX AAW01111;
AC AAW01111;
DT 01-JAN-1997 (first entry)
XX HER-2/neu protein.
DE
XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 676..1255
FT /label= Intracellular domain
FT /note= "claimed domain, useful for immunisation"
XX
PN W09630514-A1.
XX
XX
PD 03-OCT-1996.
XX
PF 28-MAR-1996; 96WO-US01689.
XX
PR 31-MAR-1995; 95US-0414417.
XX
XX (UNIW) UNIV WASHINGTON.
XX
XX Cheever MA, Disis ML;
XX
XX WPI; 1996-455361/45.
XX N-PSDB; AAT40739.
XX
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
XX Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 97.5%; Score 6631; DB 17; Length 1255;
Best Local Similarity 97.6%; Pred No. 0;
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPFGAASTQVCTGTDMLKRLPASPTHLDMLRHLQSCVQVQGNL 60
Db 1 MELAALCRWGLLLALLPFGAASTQVCTGTDMLKRLPASPTHLDMLRHLQSCVQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIANQVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIANQVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLIQIRNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLIQIRNPOLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSILTRVCAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSILTRVCAGGCARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIT 360
Db 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIT 360
Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQVFTLEBEITGYLISAWPDSLP 420
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQVFTLEBEITGYLISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGSLRSLRGLSLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGSLRSLRGLSLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFRQYIKANSKFITGITELECVGEGGLACHQLCARGHCWGPQTCCVNCQSOFLRGQEC 540
Db 481 PWDQLFRPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPQTCCVNCQSOFLRGQEC 540
Qy 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600
Qy 601 PSYKVPDLSYMPIWKFPDEGACQPCINCHTSCVDLDDKCPAQRASPLTISIVAVG 660
Db 601 PSYKVPDLSYMPIWKFPDEGACQPCINCHTSCVDLDDKCPAQRASPLTISIVAVG 660
Qy 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAOIRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAOIRILKETEL 720
Qy 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 780
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTPMPYGCGLDHRVNRGRGLGSQDLNWCMIAGMSYLEVDV 840
Db 781 YVSRLLGICLTSTVQLVTPMPYGCGLDHRVNRGRGLGSQDLNWCMIAGMSYLEVDV 840

QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYIMVKCWM 960
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYIMVKCWM 960
QY 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
DB 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
QY 1021 EYLVVQOQFFCDPAPGACGMVHRRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080
DB 1021 EYLVVQOQFFCDPAPGACGMVHRRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSLPQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAAGLQSLPHTDPSLPQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1141 NQDVRPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVYKDVFAFGGAVENPEYLTQP 1200
DB 1141 NQDVRPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVYKDVFAFGGAVENPEYLTQP 1200
QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
DB 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 6
AAW92406
ID AAW92406 standard; Protein; 1255 AA.
XX
AC AAW92406;
XX
DT 21-APR-1999 (first entry)
XX
DE Human HER-2/neu oncogene protein.
XX
KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW malignancy; treatment; tumour.
XX
OS Homo sapiens.
XX
FH Key
FT Region 676..1255
FT /note= "region which elicits immune response"
XX
PN US5869445-A.
XX
PD 09-FEB-1999.
XX
PF 01-APR-1996; 96US-0625101.
XX
PR 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Cheever MA, Disis ML;
XX
DR WPI; 1999-152835/13.
DR N-PSDB; AAX01912.
XX
PT Use of HER-2/neu polypeptides - for eliciting an immune response to
PT an HER-2/neu associated malignancy, particularly for treating or
PT preventing tumours
XX
PS Claim 3; Column 31-38; 26pp; English.
XX
CC This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.
XX
SQ Sequence 1255 AA;
Query Match 97.5%; Score 6631; DB 20; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPPGAASQTCTGTDMLKRLPASPTHLDMLRHLQGCQVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASQTCTGTDMLKRLPASPTHLDMLRHLQGCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEQVGYVLIHNOVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEQVGYVLIHNOVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIORNPOLCYODTILWKDIFHKNOLA 180
DB 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIORNPOLCYODTILWKDIFHKNOLA 180
QY 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKPLPTDCCHEQC 240
DB 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKPLPTDCCHEQC 240
QY 241 AAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGHQYIKANSKFIT 360
DB 301 YNYLSTDVSGCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGHQYIKANSKFIT 360
QY 361 ELEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLQVPEETLEEITGYLISAWPSL 420
DB 361 IQSFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLQVPEETLEEITGYLISAWPSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRSLGSLALIHNNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRSLGSLALIHNNTHLCFVHTV 480
QY 481 PWDQLFRQYIKANSKFITIGITELECVGEGLACHOLCARGHCWGPGTQCVCNCSQFLRG 540
DB 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVCNCSQFLRG 540
QY 541 VEBCRVLOGLPREYVVARHCLPCHPECPONGSVTCFPEADOCVACAHYKDPFCVARC 600
DB 541 VEBCRVLOGLPREYVVARHCLPCHPECPONGSVTCFPEADOCVACAHYKDPFCVARC 600
QY 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQASPLTSIVSAVVG 660
DB 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQASPLTSIVSAVVG 660
QY 661 ILLVVLGVVFGILIKRROOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
DB 661 ILLVVLGVVFGILIKRROOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEITLDEAYVMAGVSP 780
DB 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEITLDEAYVMAGVSP 780
QY 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSGDQLLNWCQIAKMSYLEDVR 840
DB 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSGDQLLNWCQIAKMSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900

Db 961 IDSECRPRFELVSEFSRMARDQRFVVIQNEGLGPASPLDSTFYRSLLEDDMDGLVDA 1020
QY 1021 EYLVPQQGFCPPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPQQGFCPPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQDVPAPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVVDVFAFGGAVENPEYLTPO 1200
Db 1141 NQDVPAPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVVVDVFAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
Db 1201 GGAAPQHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 8

AAV84780
ID AAY84780 standard; Protein; 1255 AA.

XX AAY84780;

XX 08-AUG-2000 (first entry)

XX Amino acid sequence of the SPLICE erbb-2 receptor protein.

XX SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.

XX Homo sapiens.

XX WO200020579-A1.

XX 13-APR-2000.

XX 01-OCT-1999; 99WO-CA00912.

XX 02-OCT-1998; 98US-0165192.

XX (UYMC-) UNIV MCMASTER.

XX Muller WJ, Siegel PM;

XX WPI; 2000-303768/26.

XX N-PSDB; AAA14812.

PT Nucleic acid encoding an erbb 2 receptor protein designated SPLICE
PT erbb-2, inhibitors of the protein are useful for treatment of cancer -

PS Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPLICE erbb-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbb-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbb-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbb-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbb-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.

XX Sequence 1255 AA;

Query Match 97.5%; Score 6631; DB 21; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;
QY 1 METAAALCRWGLLALLPPGAASTQVCTGTDMLKRLPASPTHLDMLRHLVGGCVVQGNL 60
Db 1 METAAALCRWGLLALLPPGAASTQVCTGTDMLKRLPASPTHLDMLRHLVGGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIEVQGYVLIHAHQVROVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIEVQGYVLIHAHQVROVPLQRLRIVRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLREQLRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNOLA 180
Db 121 DPLNNTPTVTGASPGGLREQLRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNOLA 180
QY 181 LTLIDNRSRACHPCSPMCKSGSCWGESSEDCOSLTRTVCAGGCARCKGPLPTCCHEQC 240
Db 181 LTLIDNRSRACHPCSPMCKSGSCWGESSEDCOSLTRTVCAGGCARCKGPLPTCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSHGI CELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSHGI CELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLHNOEVTABDGTQRCCKSPCARVCYGLGMOYIKANSKFQIT 360
Db 301 YNYLSTDVGSCTLVCPLHNOEVTABDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
QY 361 ELRFACKKIFGSLAFELPESFDGDPASNTAPLOEQLOVFETLEEITGYLISAWPSLDP 420
Db 361 IQEFACKKIFGSLAFELPESFDGDPASNTAPLOEQLOVFETLEEITGYLISAWPSLDP 420
QY 421 DLSVFQNLQVIRGRILHNGAYS LTLQGLGISWLGRLSLRELGSGLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYS LTLQGLGISWLGRLSLRELGSGLALIHNNTHLCFVHTV 480
QY 481 PWDQLFRQYIKANSKFQITTELECVGEGLAACHQLCARGHCWGPGPTQCVNCSQFLRGQBC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAACHQLCARGHCWGPGPTQCVNCSQFLRGQBC 540
QY 541 VEECRVQLGLPREYVNAHCLPCHPECQONGSVTCFGEADQCAHAHYKPPFCVARC 600
Db 541 VEECRVQLGLPREYVNAHCLPCHPECQONGSVTCFGEADQCAHAHYKPPFCVARC 600
QY 601 PSGVKPDLSPYMPKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPYMPKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKYMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
QY 721 RKVKVTLGSGAFGTGYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
Db 721 RKVKVTLGSGAFGTGYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVTOLMPYGCILLDHVRENRGLGSDLLNWCQIAKGSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTOLMPYGCILLDHVRENRGLGSDLLNWCQIAKGSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETBYHADGGKVPKKNMALESLIRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETBYHADGGKVPKKNMALESLIRRRFT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVKCWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVKCWM 960
QY 961 IDSECRPRELVSEFSRMAROPQRFVVIQNEGLGPASPLDSTFYRSLLEDDMDGLVDA 1020
Db 961 IDSECRPRELVSEFSRMAROPQRFVVIQNEGLGPASPLDSTFYRSLLEDDMDGLVDA 1020
QY 1021 EYLVPQQGFCPPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080

Db 1021 EYLVPQQGFCPPAPGAGMVRHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGKQLSLTHDPSPLQRYSDPTVPLPSETDGVYAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGKQLSLTHDPSPLQRYSDPTVPLPSETDGVYAPLTCSPQPEYV 1140
Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVVKOVFAFGGAVENPEYLTPO 1200
Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERPKTSLSPGKNGVVVKOVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPADNLYWDDPPPERGAPSTPKGTPTAENPEYLGLDVY 1255
Db 1201 GGAAPQHPHPPAFSPADNLYWDDPPPERGAPSTPKGTPTAENPEYLGLDVY 1255

RESULT 9

AAB85458
ID AAB85458 standard; Protein; 1255 AA.

XX AC AAB85458;

XX XX 25-SEP-2001 (first entry)

DT Human HER-2/neu protein.

DE Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
XX KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.

XX OS Homo sapiens.

XX PN WO200153463-A2.

XX XX 26-JUL-2001.

XX XX 19-JAN-2001; 2001WO-US01850.

XX XX 21-JAN-2000; 2000US-0177545.

XX XX (CORI-) CORIXA CORP.

XX XX Cheever MA, Hand-Zimmermann S;

XX DR WPI; 2001-476112/51.

XX DR N-PSDB; AAH23392.

XX PT New antigen-presenting cells, useful as vaccines for eliciting or
XX PT enhancing an immune response to HER-2/neu protein, particularly useful
XX PT for treating or preventing cancer, e.g. breast cancer -

XX PS Claim 2; Page 41-46; 49pp; English.

XX CC The invention provides an isolated antigen-presenting cell, which
XX CC expresses at least an immunogenic portion of a polypeptide that produces
XX CC an immune response to HER-2/neu protein. The antigen-presenting cells are
XX CC useful as vaccines for eliciting or enhancing an immune response to
XX CC HER-2/neu protein, particularly in treating or preventing malignancies in
XX CC which the HER-2/neu oncogene is associated. Specifically, these are
XX CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
XX CC colon, lung or prostate cancers. The present sequence represents
XX CC the human HER-2/neu protein (also known as p185 or c-erbB2).

XX SQ Sequence 1255 AA;

Query Match 97.5%; Score 6631; DB 22; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASQTCTGTDKMLRPLASPETHLDMLRHLVYQGCVVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASQTCTGTDKMLRPLASPETHLDMLRHLVYQGCVVQGNL 60

Qy 61 ELTYLPTNASLFLQDIEVQGYVLIHNRQVRQVPLQRLIRVRCQTQLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLFLQDIEVQGYVLIHNRQVRQVPLQRLIRVRCQTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVTCASPGGLRLQLRSLTEILKGGVLIQRNPOLCYODTILWKDI FHKNNOLA 180
Db 121 DPLNNTTPTVTCASPGGLRLQLRSLTEILKGGVLIQRNPOLCYODTILWKDI FHKNNOLA 180
Qy 181 LTLTIDNRSRACHPCSPMKGSRGCSWGSSEDCQSLTRTVCAAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLTIDNRSRACHPCSPMKGSRGCSWGSSEDCQSLTRTVCAAGGCARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSKSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSKSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHNOEVTAECDTQCEKSKPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPLNHNOEVTAECDTQCEKSKPCARVCYGLGMEHLREVRVAVTSAN 360
Qy 361 ELSEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTLEETTCGLYLSAWPDSL 420
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTLEETTCGLYLSAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTQGLIGISWLGSLRSLRGLGSLALIHNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLIGISWLGSLRSLRGLGSLALIHNTHLCFVHTV 480
Qy 481 PWDQLPROYIKANSKFIGITELECVGRGLACHOLCARGHCGWGPPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECEVGEGLACHQJCARGHCGWGPPTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSTYMPKPFDEBAGACQPCINCTHSCVDLDDKCPAEQASPLTTSIVSAYVG 660
Db 601 PSGVKPDLSTYMPKPFDEBAGACQPCINCTHSCVDLDDKCPAEQASPLTTSIVSAYVG 660
Qy 661 ILLVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEAYVMAGVSP 780
Db 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEAYVMAGVSP 780
Qy 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHVRENRGRGSDLLNWCQIAKMSYLEVDV 840
Db 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHVRENRGRGSDLLNWCQIAKMSYLEVDV 840
Qy 841 LVHRDLAARNVLKSPNHVKITDGLARLADIDETEHADGKVPKIKWMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLADIDETEHADGKVPKIKWMALESILRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYIMVMKCM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYIMVMKCM 960
Qy 961 IDSECRPRFRELVSFESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGDLVDA 1020
Db 961 IDSECRPRFRELVSFESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGDLVDA 1020
Qy 1021 EYLVPOQGFPCDPPAGAGMVRHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPOQGFPCDPPAGAGMVRHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGKQLSLTHDPSPLQRYSDPTVPLPSETDGVYAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGKQLSLTHDPSPLQRYSDPTVPLPSETDGVYAPLTCSPQPEYV 1140
Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVVKOVFAFGGAVENPEYLTPO 1200

Db 1141 NQDVRPQPSREGPLPAARPAATLERPKTLSPGKNGVUVKQVAFSGGAVENPEYLTPQ 1200
AA888267
Qy 1201 GGAAPQHPPPAPSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPPPAPSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 10
AA888267
ID AAG88267 standard; Protein; 1255 AA.
XX
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
XX
XX WO200141787-A1.
XX
XX 14-JUN-2001.
XX
XX 11-DEC-2000; 2000WO-US33591.
XX
XX 10-DEC-1999; 99US-0458299.
XX
XX (EPIM-) EPIMMUNE INC.
XX
XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
XX Keogh E;
XX WPI; 2001-374995/39.
XX
XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
XX cellular immune responses for the prevention and treatment of cancer -
XX
XX Disclosure; Page 15; 199pp; English.
XX
XX The present invention describes isolated prepared HER2/neu epitopes (I).
XX Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
XX culture in vitro and binds to a complex of an epitope (I), bound to a
XX human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
XX and a second epitope and the peptide is less than 50 contiguous amino
XX acids that have 100% identity with a native peptide sequence of HER2/neu;
XX (3) a vaccine composition (III) comprising (II) and a pharmaceutical
XX excipient; (4) an isolated nucleic acid encoding a peptide comprising
XX (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
XX and immunostimulant activities, and can be used in vaccines. (I), (II)
XX and (III) are useful for inducing cellular immune responses for the
XX prevention and treatment of cancer. (I) and (II) are useful for
XX monitoring or evaluating an immune response to a tumour-associated
XX antigen when incubated with a T lymphocyte sample from a patient and
XX detecting the presence of bound T lymphocyte to (I) or (II). Epitope
XX based vaccines mean that immunosuppressive epitopes that may be present
XX in whole antigens may be avoided. Selected epitopes may be combined to
XX enhance immunogenicity. The possible pathological side effects caused by
XX infectious agents or whole protein antigen is eliminated. The vaccine
XX provides the ability to direct and focus an immune response to multiple
XX selected antigens from the same pathogen. Epitope-based anti-tumour
XX vaccines provides the opportunity to combine epitopes derived from
XX multiple tumour-associated molecules addressing the problem of tumour-
XX tumour variability and reducing the likelihood of tumour escape due to
XX antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
XX the exemplification of the present invention.
XX
XX Sequence 1255 AA;

Query Match 97.5%; Score 6631; DB 22; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPFGAASCTGCTDMKRLPASPETHLDMRLHLYGCCVQGNL 60
Db 1 MELAALCRWGLLALLPFGAASCTGCTDMKRLPASPETHLDMRLHLYGCCVQGNL 60

Qy 61 ELYLPTNASLSFLQDIOEVQGVYLIHNNQVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Db 61 ELYLPTNASLSFLQDIOEVQGVYLIHNNQVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120

Qy 121 DPLNNTTPTVGTASPGGLRELQSLTEILKGGVLIQRNPOLCYODTILWKDI FHKNNOLA 180
Db 121 DPLNNTTPTVGTASPGGLRELQSLTEILKGGVLIQRNPOLCYODTILWKDI FHKNNOLA 180

Qy 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGGCARCKGPLTDCCHQEC 240
Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGGCARCKGPLTDCCHQEC 240

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGHQYI KANSKFIT 360
Db 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGHQYI KANSKFIT 360

Qy 361 ELEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVPTLEETIGLYLISAWPDSLP 420
Db 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVPTLEETIGLYLISAWPDSLP 420

Qy 421 DLSVFQNLQVIRGRIHLNGAYSILTLOGLGISWLGSLRLSELGSLALIHNNTHLCFVHT 480
Db 421 DLSVFQNLQVIRGRIHLNGAYSILTLOGLGISWLGSLRLSELGSLALIHNNTHLCFVHT 480

Qy 481 PWDQLFRQYIKANSKFITTELECEVGEGLACHQICARGHCWGPGPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNQHALLHTANRPEDECEVGEGLACHQICARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy 541 VECRVLQGLPREYVNAHCLPCHPECPONGSVTCFPGPADQCVACAHYKDPFCVARC 600
Db 541 VECRVLQGLPREYVNAHCLPCHPECPONGSVTCFPGPADQCVACAHYKDPFCVARC 600

Qy 601 PSGVKPDLSTYMPKPFDEEGACQPCPINCTHSCVDLDDKGCQAERASPLTISVAVG 660
Db 601 PSGVKPDLSTYMPKPFDEEGACQPCPINCTHSCVDLDDKGCQAERASPLTISVAVG 660

Qy 661 ILLVVLGVVFGILIKRQOKIRKYTNRRLLQETELVEPLTPSGAMPNQAOMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYTNRRLLQETELVEPLTPSGAMPNQAOMRILKETEL 720

Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVMAGVSP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVMAGVSP 780

Qy 781 YVSRLLGICLTSTVQLVTLQMPYGCILLDHYENRGLGSDLLNWCQIAGMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYGCILLDHYENRGLGSDLLNWCQIAGMSYLEDVR 840

Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRFT 900

Qy 901 HQSDVMSYGVYVWELMTFGAKPYDGI PAREIPDLEKGERLPQPPICTIDVYMIWVKCW 960
Db 901 HQSDVMSYGVYVWELMTFGAKPYDGI PAREIPDLEKGERLPQPPICTIDVYMIWVKCW 960

Qy 961 IDSECPREFELVSEFSRMARDPQRFVWQNEQDLGPASPLDSTFYRSLLEDMDGLVDA 1020
Db 961 IDSECPREFELVSEFSRMARDPQRFVWQNEQDLGPASPLDSTFYRSLLEDMDGLVDA 1020

Qy 1021 EYLVFPQQGFCCDDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Db 1021 EYLVFPQQGFCCDDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080

Db 1021 EYLVPOQGFPCDPAPACAGMWHRRSSSTRSGGDLTLGLPSEBEEAPRSLAPSEG 1080
QY 1081 AGSDVDFDGLMGAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVDFDGLMGAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1141 NQPDVVRPQPPSPREGPLPAARPAAGATLIERAKTSLSPGKNGVVKOVFAFGGAVENPEYLTQP 1200
Db 1141 NQPDVVRPQPPSPREGPLPAARPAAGATLIERPKTLPFGKNGVVKOVFAFGGAVENPEYLTQP 1200
QY 1201 GGAAPQHPHPPAPSPADNLYYWDQDPPERCAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPHPPAPSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 11

AAE24067
ID AAE24067 standard; Protein; 1255 AA.
XX AAE24067;
XX 23-SEP-2002 (first entry)
XX Human Her-2 protein.
XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
XX Homo sapiens.
OS Homo sapiens.
XX WO200222636-A1.
XX 21-MAR-2002.
XX 12-SEP-2001; 2001WO-US28572.
XX 15-SEP-2000; 2000US-0663834.
XX (ISIS-) ISIS PHARM INC.
XX Bennett CF, Cowse LM;
XX WPI; 2002-471192/50.
DR N-PSDB; AAD38904.
XX Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX Example 13; Page 95-107; 116pp; English.

XX The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.

XX Sequence 1255 AA;

Query Match 97.5%; Score 6631; DB 23; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASOVCTGTDMLRLPASPEHLDMLRHLVQGCVOVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASOVCTGTDMLRLPASPEHLDMLRHLVQGCVOVQGNL 60

QY 61 ELTYLPTNASLFLQDIQEVGYVLI AHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIQEVGYVLI AHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVTCSPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDI FHKNNQLA 180
Db 121 DPLNNTTPTVTCSPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDI FHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPMCKGSRGCMGSESSDCQSLTRTVCAAGGCARCKPGLPTDCCHQC 240
Db 181 LTLIDTNRSRACHPCSPMCKGSRGCMGSESSDCQSLTRTVCAAGGCARCKPGLPTDCCHQC 240
QY 241 AAGCTGPKHSKSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Db 241 AAGCTGPKHSKSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRVAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRVAVTSAN 360
QY 361 ELREFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFEETLEETGTYLVIISAWPDSLP 420
Db 361 IQEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFEETLEETGTYLVIISAWPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLISWGLSLRLSLRELGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLISWGLSLRLSLRELGLALIHNTLHLCFVHTV 480
QY 481 PWDOLFROYIKANSKFKITTELECVGEGGLACHQICARGHCWGPGPTQCVCNCSQFLRGQEC 540
Db 481 PWDOLFPRNHQALLHTANRPEDECEVGEGLACHQICARGHCWGPGPTQCVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREVYNARHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREVYNARHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 601 PSGVKPDLSPYPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQASPLTISVAVVG 660
Db 601 PSGVKPDLSPYPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQASPLTISVAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
QY 721 RKVKVLGSGAGFTYKGIWIPDGENVKI PVAIKVLRNTPSKANKETILDEAYVNAVGVSP 780
Db 721 RKVKVLGSGAGFTYKGIWIPDGENVKI PVAIKVLRNTPSKANKETILDEAYVNAVGVSP 780
QY 781 YVSRLLGICLTSTVQLVTQLMPYGLLDHVRNRRGRGLSQDGLLNCWCMQIAKGMVSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQLMPYGLLDHVRNRRGRGLSQDGLLNCWCMQIAKGMVSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPIKWMALLESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPIKWMALLESILRRRFT 900
QY 901 HQSDVMSYGVTVVLMETFGAKPDYGIIPARI PDLEKGERLPPOPICTIDVYIMLVKCMW 960
Db 901 HQSDVMSYGVTVVLMETFGAKPDYGIIPARI PDLEKGERLPPOPICTIDVYIMLVKCMW 960
QY 961 IDSECRPRPRELVSEFSRMDARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDGDLVDA 1020
Db 961 IDSECRPRPRELVSEFSRMDARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDGDLVDA 1020
QY 1021 EYLVPOQGFPCDPAPACAGMWHRRSSSTRSGGDLTLGLPSEBEEAPRSLAPSEG 1080
Db 1021 EYLVPOQGFPCDPAPACAGMWHRRSSSTRSGGDLTLGLPSEBEEAPRSLAPSEG 1080
QY 1081 AGSDVDFDGLMGAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVDFDGLMGAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1141 NQPDVVRPQPPSPREGPLPAARPAAGATLIERAKTSLSPGKNGVVKOVFAFGGAVENPEYLTQP 1200

Db 1141 NQDVRPQPSPREGPLPAARPAAGATLERPKTLSPGKNGVGVKOVFAFGGAVENPEYLTPQ 1200
QY 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSFKGTPTTAENPEYLGDDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSFKGTPTTAENPEYLGDDVPV 1255

RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX
AC AAE20479;
XX
DT 01-JUL-2002 (first entry)
XX
DE Human Her-2/neu protein.
XX
KW Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
XX
XX WO200214503-A2.
XX
XX 21-FEB-2002.
XX
XX 14-AUG-2001; 2001WO-US41733.
XX
XX 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-236428P.
PR 21-FEB-2001; 2001US-270520P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedvick TS;
XX
XX WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer -
XX
XX Disclosure; Page 114-117; 129pp; English.
XX
XX The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.
XX
SQ Sequence 1255 AA;

Query Match 97.5%; Score 6631; DB 23; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;

QY 1 MELAAALCRWGLLALLPPGAASSTQVCTGTDMKRLPAS PETHLDMLRHLVGGCVQVQGNL 60
Db 1 MELAAALCRWGLLALLPPGAASSTQVCTGTDMKRLPAS PETHLDMLRHLVGGCVQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVROVPLQRLRIVRGTRQTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVROVPLQRLRIVRGTRQTLFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLREQLRSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNOLA 180
Db 121 DPLNNTTPVTGASPGGLREQLRSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNOLA 180
QY 181 LTLIDTNRSRACHPCSPMKSGSCWGSSEDCSLTRTVCAGGCARCKPLPTDCCHQEC 240
Db 181 LTLIDTNRSRACHPCSPMKSGSCWGSSEDCSLTRTVCAGGCARCKPLPTDCCHQEC 240
QY 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHMQEVTAE DGTQRCCKSPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHMQEVTAE DGTQRCCKSPCARVCYGLGMEHLREVRVAVTSAN 360
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLISAWPSLP 420
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLISAWPSLP 420
QY 421 DLSVFQNLQVIRGIRILHNGAYSLTLQGLIGISWLGSLRSLRELGSGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGIRILHNGAYSLTLQGLIGISWLGSLRSLRELGSGLALIHNTLHLCFVHTV 480
QY 481 PWDQLFRQYIKANSKFIGITELECVGEGLA CHOLCARGHCGWGPGTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLA CHOLCARGHCGWGPGTQCVNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECOPONGSVTCFGEADOCVACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECOPONGSVTCFGEADOCVACAHYKDPFPCVARC 600
QY 601 PSGVKPDLSPYPIWKFPDBEGACQPCPINCTHSCVDLDDKGCAPABQASPLTSTVSVAWG 660
Db 601 PSGVKPDLSPYPIWKFPDBEGACQPCPINCTHSCVDLDDKGCAPABQASPLTSTVSVAWG 660
QY 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVTOLMPYGCLLDHVRENRGRIGSDLLNWCQIAGKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTOLMPYGCLLDHVRENRGRIGSDLLNWCQIAGKMSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWMALESILRRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVWKCM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVWKCM 960
QY 961 IDSECRPRELVSEFSRMRAPQREVIQNE DLGPASPLDSTFYRSLLEDDDMGLVDA 1020
Db 961 IDSECRPRELVSEFSRMRAPQREVIQNE DLGPASPLDSTFYRSLLEDDDMGLVDA 1020
QY 1021 EBYLVPQGGFFCPCDPAPGAGGMVHRRHSSTSRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Db 1021 EBYLVPQGGFFCPCDPAPGAGGMVHRRHSSTSRSGGDLTLGLEPSEEEAPRPLAPSEG 1080

Db 1021 EYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1141 NQDVRPQPPSPREGPLPAARPAAGATLRAKTLSPGKNGVVKOVFAFGGAVENPEYLTTPQ 1200
Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLRAKTLSPGKNGVVKOVFAFGGAVENPEYLTTPQ 1200
QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERCAPSTFKGPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERCAPSTFKGPTAENPEYLGLDVVP 1255

RESULT 13
AAM51143
ID AAM51143 standard; Protein; 1255 AA.
XX AC AAM51143;
XX DT 17-JUN-2002 (first entry)
XX DE Human Her-2/neu oncogene-encoded p185 glycoprotein.
XX KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
XX KW tyrosine kinase; receptor; c-erbB2; gene therapy.
XX OS Homo sapiens.

PH Key Location/Qualifiers
FT Domain 1..653 /note= "extracellular domain"
FT FT 676..1255 /note= "intracellular domain"
FT FT 990..1255 /note= "phosphorylation domain"
XX WO200212341-A2.
XX PN 14-FEB-2002.
XX PD 03-AUG-2001; 2001WO-US24283.
XX PF 03-AUG-2000; 2000US-0632507.
XX PR (CORI-) CORIXA CORP.
XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX PI Cheever MA, Gheysen D;
XX DR WPI; 2002-241743/29.
XX DR N-PSDB; ABA92250.
XX PT Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT or enhancing an immune response to the protein, has Her-2/neu
PT extracellular domain fused to Her-2/neu intracellular or
PT phosphorylation domain
XX Claim 68; Fig 7; 14pp; English.

CC The present sequence is that of human Her-2/neu (p185 glycoprotein
CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
CC in a variety of cancers, including breast, ovarian, colon, lung and
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
CC family of receptor-like glycoproteins. It comprises an extracellular
CC domain with homology to the epidermal growth factor receptor
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC intracellular domain that also shows homology to EGFR. Its
CC overexpression correlates with a poor prognosis in breast and
CC ovarian cancers. The invention provides Her-2/neu fusion
CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its DeltaPD fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.
XX
SQ Sequence 1255 AA;
Query Match 97.5%; Score 6631; DB 23; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPEHLDMLRHLVQGCQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPEHLDMLRHLVQGCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGVYVIAHNOVROVPLQRLRIVRGTTQQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGVYVIAHNOVROVPLQRLRIVRGTTQQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTTLWKDIFHKNNOLA 180
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTTLWKDIFHKNNOLA 180
QY 181 LTLIDTNRSPACHPCSPMCKGSRGWSESDCQSLTRTVAGGCARCKGPIPTDCCHRCQ 240
Db 181 LTLIDTNRSPACHPCSPMCKGSRGWSESDCQSLTRTVAGGCARCKGPIPTDCCHRCQ 240
QY 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNHOEVTAEDGTORCEKSKPCARVCYGLGQYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPLNHOEVTAEDGTORCEKSKPCARVCYGLGMEHLREAVTSAN 360
QY 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLEETGTYLISAWPDSLP 420
Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLEETGTYLISAWPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWLGRLSLRELGLSLIHHNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWLGRLSLRELGLSLIHHNTHLCFVHTV 480
QY 481 PWDLFRQYIKANSKFITITELCVGEGGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Db 481 PWDLFRQYIKANSKFITITELCVGEGGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 601 PSGVKPDLSPYMPITWKPFDDEGACQPCPINCTHSCVDLDDKGCPEQRASPLTISVAVVG 660
Db 601 PSGVKPDLSPYMPITWKPFDDEGACQPCPINCTHSCVDLDDKGCPEQRASPLTISVAVVG 660
QY 661 ILLVVLGVVFGILIKERQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRLIKETEL 720
Db 661 ILLVVLGVVFGILIKERQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRLIKETEL 720
QY 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDBAYVMAGVSP 780
Db 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDBAYVMAGVSP 780

QY 781 YVSRLLGICLTSTVQLVTLMPYGCCLLDHVRNRRGLSGDILLNWCQIAKMSYLEDDR 840
 DB 781 YVSRLLGICLTSTVQLVTLMPYGCCLLDHVRNRRGLSGDILLNWCQIAKMSYLEDDR 840
 QY 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRRFT 900
 DB 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRRFT 900
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
 QY 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
 DB 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
 QY 1021 EYLVPQGGFFCPDPAPGAGMWHRRSSSTRSGGDLTLGLSEPEEAPRSPAPSEG 1080
 DB 1021 EYLVPQGGFFCPDPAPGAGMWHRRSSSTRSGGDLTLGLSEPEEAPRSPAPSEG 1080
 QY 1081 AGSDVFDGDLGMAAKGLSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 DB 1081 AGSDVFDGDLGMAAKGLSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 QY 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTQ 1200
 DB 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTQ 1200
 QY 1201 GGAAPQHPHPPAFSPAFDNLVYDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
 DB 1201 GGAAPQHPHPPAFSPAFDNLVYDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 14

AAU77114
 ID AAU77114 standard; Protein; 1255 AA.

XX AC AAU77114;

XX DT 05-JUN-2002 (first entry)

XX DE Human Her-2/neu polypeptide.

XX KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
 KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
 KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
 XX Hodgkin's lymphoma; T cell therapy.

XX OS Homo sapiens.

XX PN WO200213847-A2.

XX PD 21-FEB-2002.

XX PF 13-AUG-2001; 2001WO-US25408.

XX PR 14-AUG-2000; 2000US-0638280.

XX PR 28-SEP-2000; 2000US-0675904.

XX XX (CORI-) CORIXA CORP.

XX XX Gaiger A, Cheever MA, Hand-zimmermann S;

XX DR WPI; 2002-280741/32.

XX DR N-PSDB; ABK10730.

XX XX Inhibiting haematological malignancy development by administering
 PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
 PT encoding the polypeptide, or antigen presenting cells expressing the
 XX polypeptide
 PS Disclosure; Page 71-74; 74pp; English.

CC The invention relates to a method for inhibiting development of
 CC haematological malignancy in a patient by administering a polypeptide
 CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
 CC encoding the polypeptide. Antigen presenting cells that express the
 CC protein can also be administered. The sequences are used for inhibiting
 CC development of haematological malignancy such as acute myelogenous
 CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
 CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
 CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
 XX

SQ Sequence 1255 AA;

Query Match 97.5%; Score 6631; DB 23; Length 1255;
 Best Local Similarity 97.6%; Pred. No. 0;
 Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASQVCTGTDMKURLPASPTHLDMLRHLVGGCCVQVQNL 60

DB 1 MELAALCRWGLLLALLPPGAASQVCTGTDMKURLPASPTHLDMLRHLVGGCCVQVQNL 60

QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIHAHQVQVPLQRLIRVRGTQLFEDNYALAVLDNG 120

DB 61 ELTYLPTNASLSFLQDIQEVQGYVLIHAHQVQVPLQRLIRVRGTQLFEDNYALAVLDNG 120

QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNQOLA 180

DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNQOLA 180

QY 181 LTLIDTNRSRACHPCSPMKCGSRGWESSEDCSLTRTVCCAGCARCKGLPTDCCHQC 240

DB 181 LTLIDTNRSRACHPCSPMKCGSRGWESSEDCSLTRTVCCAGCARCKGLPTDCCHQC 240

QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

QY 301 YNYLSTDVSGSCTLVCPLNQEVTAEDGTQCEKSKPCARVCYGLGHWYIKANSKFIT 360

DB 301 YNYLSTDVSGSCTLVCPLNQEVTAEDGTQCEKSKPCARVCYGLGHWYIKANSKFIT 360

QY 361 ELDFAGCKITFGSLAFIPESFGDPPASNTAPLOPEOLQVPEETEEITGYLYISAWPSLP 420

DB 361 IQDFAGCKITFGSLAFIPESFGDPPASNTAPLOPEOLQVPEETEEITGYLYISAWPSLP 420

QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLSLRSLRELGLSLALIHNTHLCHFVHT 480

DB 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLSLRSLRELGLSLALIHNTHLCHFVHT 480

QY 481 PWDQLPRQYIKANSKFITTELECVGEGLAHQLCARGHCWGPQTCVNCSPQLRQEC 540

DB 481 PWDQLPRQYIKANSKFITTELECVGEGLAHQLCARGHCWGPQTCVNCSPQLRQEC 540

QY 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVCAHYKDPFPCVARC 600

DB 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVCAHYKDPFPCVARC 600

QY 601 PSGVKPDLSYMPTWKPPDEEGACQPCINCTHSCVDLDDKGCPEARASPLTISVSAVVG 660

DB 601 PSGVKPDLSYMPTWKPPDEEGACQPCINCTHSCVDLDDKGCPEARASPLTISVSAVVG 660

QY 661 ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

DB 661 ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

QY 721 RKVKVLGSGAGFVTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYMAGVGP 780

DB 721 RKVKVLGSGAGFVTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYMAGVGP 780

QY 781 YVSRLLGICLTSTVQLVTLMPYGCCLLDHVRNRRGLSGDILLNWCQIAKMSYLEDDR 840

DB 781 YVSRLLGICLTSTVQLVTLMPYGCCLLDHVRNRRGLSGDILLNWCQIAKMSYLEDDR 840

QY 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRRFT 900

```
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGKVPKWMMALESILRRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIMVWKCM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIMVWKCM 960
Qy 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
Qy 1021 EYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Db 1021 EYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGKGLQSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGKGLQSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLT PQ 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERPKTLPSPGKNGVVKDVFAFGGAVENPEYLT PQ 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVPV 1255
```

RESULT 15

AA39568 ID AA39568 standard; Protein; 1433 AA.

XX AC AA39568;

XX DT 07-FEB-1994 (first entry)

XX DE Sequence of c-erbB-2 tumour antigen.

XX KW Tumour antigen; c-erbB-2; glycoprotein.

XX OS Homo sapiens.

XX PN WO9316185-A.

XX PD 19-AUG-1993.

XX PF 05-FEB-1993; 93WO-US01055.

XX PR 06-FEB-1992; 92US-0831967.

XX PA (CETU) CETUS ONCOLOGY CORP.

XX PI (CREA-) CREATIVE BIOMOLECULES INC.

XX PI Houston LL, Huston JS, Oppermann H, Ring DB;

XX DR WPI; 1993-272889/34.

XX DR N-PSDB; AAQ46083.

XX PT New single chain Fv polypeptide binding to C-erbB-2 tumour

XX PT antigen - for imaging or treating breast or ovarian cancer etc.

XX PS Disclosure; pages 48-54; 87pp; English.

XX CC c-erbB-2 refers to a protein antigen expressed on the surface of

XX CC tumour cells, such as breast and ovarian tumour cells, which is an

XX CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric

XX CC pt. of about 5.3 (see AAQ46083; AAR39568). X in AAR39583 represents

XX CC the location of a stop codon in AAQ46083.

XX SQ Sequence 1433 AA;

Query Match

Best Local Similarity

96.8%; Score 6588; DB 14; Length 1433;

97.1%; Pred. No. 0;

```
Matches 1218; Conservative 8; Mismatches 29; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPPGAASSTOCTGTDMLKRLPASPETHLDMLRLHYQGCQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTOCTGTDMLKRLPASPETHLDMLRLHYQGCQVQGNL 60
Qy 61 ELTYLPTNASLFLQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Db 61 ELTYLPTNASLFLQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Qy 121 DPLNNTTPTVCASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKQNOLA 180
Db 121 DPLNNTTPTVCASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKQNOLA 180
Qy 181 LTLTDVNRSRACHPCSPMKGSRGWSSESSEDCQSLTRTVCAAGGACRCKGPLTDCCHQOC 240
Db 181 LTLTDVNRSRACHPCSPMKGSRGWSSESSEDCQSLTRTVCAAGGACRCKGPLTDCCHQOC 240
Qy 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCITLVCPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCITLVCPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFFETLEEITGTYLISAWPDSL 420
Db 361 IOEPAGCKRIFGSLAFLPESFDGDPASNTAPLQPEHLQVFFETLEITGTYLISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYS LTLQGLISWLGSLRLSRLGSLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYS LTLQGLISWLGSLRLSRLGSLALIHNNTHLSFVHTV 480
Qy 481 PWDOLFROYIKANSKFITGITELECVGEGGLACHQICARGHCWGPGTQCWCNQSFURGQEC 540
Db 481 PWDOLFNRPHQALLHTANRPEDECVGEGGLACHQICARGHCWGPGTQCWCNQSFURGQEC 540
Qy 541 VEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPPECVARGC 600
Db 541 VEECRVLOGLPREVYNASHCLPCHPECPQNGSVTCFGEADQCACAHYKDPPECVARGC 600
Qy 601 PSYKVPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCPCAEQASPLTSIVSAVVG 660
Db 601 PSYKVPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCPCAEQASPLTSIVSAVVG 660
Qy 661 ILLVVVLGVVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKI PVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKI PVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGRGSLQDLNWCQIAKMSYLEDDR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGRGSLQDLNWCQIAKMSYLEDDR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGKVPKWMMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGKVPKWMMALESILRRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIMVWKCM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIMVWKCM 960
Qy 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
Qy 1021 EYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Db 1021 EYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
```

```
Qy 1081 AGSDVFDGDLGMGAAGKGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMGAAGKGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERPKTLPSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVTV 1255
Db 1201 GGAAPQPHPPPAFSPAFDNLYYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVTV 1255
```

Search completed: July 22, 2003, 08:41:34
Job time : 42.9774 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-465-479-12

Perfect score: 6804

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPSYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	6631	97.5	1255	1 ERB2 HUMAN	P04626 homo sapien
2	5851.5	86.0	1254	1 ERB2 MESAU	P06553 mesocricetu
3	5850	86.0	1257	1 ERB2 RAT	P06494 rattus norv
4	3119	45.8	1210	1 EGFR HUMAN	P00533 homo sapien
5	3088	45.4	1210	1 EGFR MOUSE	Q01279 mus musculu
6	2943.5	43.3	1308	1 ERB4 HUMAN	Q15303 homo sapien
7	2921	42.9	1308	1 ERB4 RAT	Q62956 rattus norv
8	2674.5	39.3	1167	1 XMRK XIPMA	P13388 xiphophoru
9	2393.5	35.2	1342	1 ERB3 HUMAN	P21860 homo sapien
10	2323.5	34.1	1339	1 ERB3 RAT	Q62799 rattus norv
11	1925	28.3	1426	1 EGFR DROME	P04412 drosophila
12	1749.5	25.7	634	1 ERBB ALV	P00534 avian leuko
13	1703	25.0	604	1 ERBB AVIER	P00535 avian eryth
14	1630	24.0	540	1 ERBB AVIEU	P11273 avian eryth
15	1575	23.1	703	1 EGFR CHICK	P13387 gallus gall
16	1429	19.0	1223	1 LT23 CAEEL	P24348 caenorhabdi
17	1142.5	16.8	245	1 ERB2 MOUSE	P70424 mus musculu
18	724	10.6	1363	1 ILPR BRALA	O02466 brachiosto
19	696	10.2	1300	1 IRR MOUSE	Q9wt14 mus musculu
20	695	10.2	1383	1 INSR RAT	P15127 rattus norv
21	695	10.2	1477	1 HTK7 HYDAT	Q25197 hydra atten
22	694.5	10.2	1372	1 INSR MOUSE	P15208 mus musculu
23	694	10.2	1382	1 INSR HUMAN	P06213 homo sapien
24	691	10.2	1297	1 IRR HUMAN	P14616 homo sapien
25	685	10.1	1607	1 MIPR LYMTS	Q25410 lymaea sta
26	682.5	10.0	1300	1 IRR CAVPO	P14617 cavia porce
27	646	9.5	1367	1 IGR1 HUMAN	P08069 homo sapien
28	629	9.2	1373	1 IGR1 MOUSE	Q60751 mus musculu
29	626	9.2	1390	1 INSR AEDAE	Q93105 aedes aegypt
30	625.5	9.2	1370	1 IGR1 RAT	P24062 rattus norv
31	616	9.1	2146	1 INSR DROME	P09208 drosophila
32	609.5	9.0	984	1 EPB1 RAT	P09759 rattus norv
33	603.5	8.9	984	1 EPB1 HUMAN	P54762 homo sapien

ALIGNMENTS

```

RESULT 1
ERB2_HUMAN
ID ERB2_HUMAN STANDARD; PRT: 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erbB-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RA "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

```

```

34 603 8.9 987 1 EPB4_HUMAN P54760 homo sapien
35 601.5 8.8 984 1 EPB1_CHICK Q07494 gallus gall
36 593 8.7 902 1 EPBB_XENLA Q91736 xenopus lae
37 587.5 8.6 985 1 EPBA_XENLA Q91571 xenopus lae
38 587 8.6 1114 1 RET_HUMAN P07949 homo sapien
39 585.5 8.6 977 1 EPA2_MOUSE Q03145 mus musculu
40 585 8.6 987 1 EPB4_MOUSE P54761 mus musculu
41 584.5 8.6 976 1 EPA2_HUMAN P29317 homo sapien
42 578.5 8.5 1053 1 FAK1_CHICK Q00944 gallus gall
43 573 8.4 757 1 HT16_HYDAT P53356 hydra atten
44 569 8.4 1068 1 FAK1_XENLA Q91738 xenopus lae
45 565.5 8.3 1052 1 FAK1_MOUSE P34152 mus musculu

```


Best Local Similarity 86.0%; Pred. No. 2.8e-305; Matches 1079; Conservative 61; Mismatches 114; Indels 1; Gaps 1;		
Qy	1	MELAAALCRWGLLIALLPAGASTOVCTGTDMKPLRPASPETHLMDLHLVGGCCVVOGNL 60
Db	1	MELAAACWGLLALLSPGASGTQVCTGTDMKPLRPASPETHLMDLHLVGGCCVVOGNL 60
Qy	61	ELTYLPTNASLFLQDIQEVQGYVLIAHNOVROPLOQLRIRVGTQFEDNYALAVLDNG 120
Db	61	ELTYLPANATLSFLQDIQEVQGYVLIASHQVHVPLQRLRIRVGTQFEDKYALAVLDNR 120
Qy	121	DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPCLQYQDTILWKDIFHNHQLA 180
Db	121	DPLDNVTATGRTPEGLRELQRLSLTEILKGGVLIQRNPCLQYQDTILWKDIFRNHQLA 180
Qy	181	LTLTDNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGGCARCKGLPTCCHEQC 240
Db	181	PVDLTDNRSRACPPCACPCKDNHCWGPASPEDCQLTGTIAPRAVPAARARLPTCCHEQC 240
Qy	241	AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
Db	241	AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTCP 300
Qy	301	YNLSTDVGSCTLVCPHNOBVTAEQGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
Db	301	YNLSTEVGSCTLVCPHNOBVTAEQGTQRCCKSKPCARVCYGLGMEHLRGARAITSAN 360
Qy	361	ELEFAGCKIFGSLAFIPESDGPASNTAPLOEQLOVFETLEETGYLYISAWPSLP 420
Db	361	IQEFAGCKIFGSLAFIPESDGPNSGIAPLTPEQLQVFETLEETGYLYISAWPSLSH 420
Qy	421	DLVSFONLQVIRGRILHNGAYSLLTQGLISWGLRSRLRELGLALIHHTHLCFCVHTV 480
Db	421	DLVSFONLVRIRGRVLDHNGAYSLLTQGLIRWGLRSRLRELGLVLIHRNTHLCFVHTV 480
Qy	481	PWDQLFRQYIKANSKFITGITELECVGEGELACHQICARGHCWGPPTQCVNCSQFLRGEC 540
Db	481	PWDQLFRPHQALLHSGNPSEEGCLKDFACYPLCAHGCWGPPTQCVNCSHFLRGEC 540
Qy	541	VEECRVQLGPREVYNARHCLPCHPEQOPONGSVTCGPEADOCVACHYKDPFCVARC 600
Db	541	VKECRVWGLPREVYNGKHCLPCHPEQOPQNSTCTGSEADQCTACPHYKDSFFCVARC 600
Qy	601	PSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGPACPAEQASPLTISVAVG 660
Db	601	PSGVKPDLSYMPIWKYPDEGCMQPCPINCTHSCVDLDERCPCPAEQASPATSIATVVG 660
Qy	661	ILLVVLGVGFIILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db	661	ILLFLVGVVVGFIILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy	721	RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 780
Db	721	RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGLSP 780
Qy	781	YVSRLLGLTSTVQLVTQMLPYGCLLDHVRNRRGLSQDQLLNWCMQIAGKMSYLEDDR 840
Db	781	YVSRLLGLTSTVQLVTQMLPYGCLLDHVRHRRGLSQDQLLNWCMQIAGKMSYLEDDR 840
Qy	841	LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
Db	841	LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
Qy	901	HQSDVWSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMTVMKWM 960
Db	901	HQSDVWSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMTVMKWM 960
Qy	961	IDSECRPRFRELVSFSEFARMARDPQFVVYIQNEDLGPASPLDSTFYRSLLEDGDLVDA 1020
Db	961	IDSECRPRFRELVSFSEFARMARDPQFVVYIQNEDLGPSSPLDSTFYRSLLEDGDLVDA 1020
Qy	1021	EYLVPQGGFFPDPAFGAGSTAHRRHSSSTRSGGELTLGMEPSPGEPSPRLAPSEG 1080
Db	1021	EYLVPQGGFFPDPAFGAGSTAHRRHSSSTRSGGELTLGMEPSPGEPSPRLAPSEG 1080

1081 AGSDVDFDGLGMGAAGKGLQSLPTHDPSPLOQYSDDPTVPLPSETDGYVAPLTCSPQPEYV 1140

1081 AGSDVDFEGLGMGATKGPQSIISPRDLSPLOQYSDDPTVPLPSETDGYVAPLTCSPQPEYV 1140

1141 NQDVRPQPPSPREGPLPAAPAGATLAKTLSPGXGVVYKVFAGFAGAVENPEYLTPQ 1200

1141 NQPEVRPQPPSPREGPLPAAPAGATLAKTLSPGXGVVYKVFAGFAGAVENPEYLTPQ 1200

1201 GGAAPQPPSPREGPLPAAPAGATLAKTLSPGXGVVYKVFAGFAGAVENPEYLTPQ 1255

1201 GGSASQPH-PPALCPAFDNLVYDQDPSERGSPPPTFEGTPTAENPEYLGLDVVP 1254

RESULT 3

ERB2_RAT

1D-ERB2_RAT STANDARD; PRT; 1257 AA.

AC P06494;

DT 01-JAN-1988 (Rel. 06, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)

DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor receptor-related protein).

GN ERB2 OR NEU.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Neuroblastoma;

RX MEDLINE=86118662; PubMed=3945311;

RA Baryman C.I., Hung M.-C., Weinberg R.A.;

RT "The neu oncogene encodes an epidermal growth factor receptor-related protein.";

RL Nature 319:226-230(1986).

RN [2]

RP SEQUENCE OF 852-905 FROM N.A.

RC TISSUE=Sciatic nerve;

RX MEDLINE=9122560; PubMed=2025425;

RA Lai C., Lemke G.;

RT "An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system.";

RL Neuron 6:691-704(1991).

RN [3]

RP STRUCTURE BY NMR OF 650-668.

RX MEDLINE=92155181; PubMed=1346763;

RA Gullick W.J., Bottomley A.C., Loftis F.J., Doak D.G., Mulvey D., Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;

RT "Three dimensional structure of the transmembrane region of the proto-oncogenic and oncogenic forms of the neu protein.";

RL EMBO J. 11:43-48(1992).

CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIREGULIN.

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.

CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

Qy 1139 YVNPQDVRRPQPPSPREGPLPAARAGATLTERAKTLSPGKNGVWVDVFAFGGAVENPEYL 1198
 Db 1141 YVNSEVQPPPLTPEGPLPVRAGATLERPKTLSPGKNGVWVDVFAFGGAVENPEYL 1200
 Qy 1199 PQGGAAPQHPHPPAFSPAFDNLVYWDQPPRGERAPPSTFKTPTAENPEYLGLDV 1255
 Db 1201 PREGTASPPHPSAFSPAFDNLVYWDQNSSEQGPPPSNFECTPTAENPEYLGLDV 1257

RESULT 4
 EGRF_HUMAN
 ID EGRF_HUMAN STANDARD; PRT; 1210 AA.
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;
 AC O00688; Q0B252; Q9H2C9; Q9GZK1; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 DE protein-tyrosine kinase ErbB-1).
 GN EGRF OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Hayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RA "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RL Nature 309:418-425(1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=95382957; PubMed=7654368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RA "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta.";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=97078686; PubMed=8918811;
 RA Reiter J.L., Maible N.J.;
 RA "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor.";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=97256547; PubMed=9103388;
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
 RA "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TEGFR) in ovarian cancer.";
 RL Gynecol. Oncol. 65:36-41(1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RC TISSUE=Placenta;
 RX MEDLINE=21100872; PubMed=11161793;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramanian S., Crossley T.D., Magnuson T.R., James C.D.,
 RA Maible N.J.;
 RA "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms.";
 RL Genomics 71:1-20(2001).
 RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
 RA Lampland A.L., Balasubramanian S., Crossley T.O., Magnuson T.R.,

RA Maible N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
 RA "Expression cloning of human EGF receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells.";
 RL Science 224:843-848(1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RA "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells.";
 RL Nature 309:806-810(1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
 RA O'Malley B.W.;
 RA "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells.";
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
 RA Waterfield M.D.;
 RA "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription.";
 RL Oncogene Res. 1:375-396(1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RA "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RL factor receptor proto-oncogene RNA synthesis.";
 RN J. Biol. Chem. 266:1746-1753(1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RA "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN [13]
 RP SEQUENCE OF 540.
 RX Kohda D.;
 RL Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RA "ATP-stimulated interaction between epidermal growth factor receptor
 RT and supercoiled DNA.";
 RL Nature 309:270-273(1984).
 RN [15]
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
 RA "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor.";
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN [16]


```

FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64;

Query Match 45.4%; Score 3088; DB 1; Length 1210;
Best Local Similarity 49.5%; Pred. No. 1.1e-157;
Matches 630; Conservative 168; Mismatches 363; Indels 112; Gaps 25;

QY 11 LLLALLPPGAA--STOVCTGTDMKLRUPASPTHLDMLRLHYLQGVVQGNLELYLPTN 68
DB 14 LTLCAAGALEBKVCQGSNRLTQGTGFEDHFLSLQRYNNCEVVLGNLEITYYQVN 73
QY 69 ASLSFLDIOEVQGYLIAHNVQVPLQRLIRVGTQLFEDNYALVLDNGDPLNNTTP 128
DB 74 YDLSFLTKIOEVAGYVLIANTVERIPLENIQIIRGNALYENTYALAILSN----- 124
QY 129 VTGASPGRLRLQRLSLEILKGVLIQRPOLCYQDTILWKDI-----FHKQNALTLI 184
DB 125 -YGTNRGLRELPRNLQELIIGAVRFNSNPILCNMDTIQWRDIVQNVFMSNMGDL--- 180
QY 185 DTRSRACHPCSPKCSRGWGESSEDCOSLTRVTCAGGCA-RCKGPLPTDCHEQCAAG 243
DB 181 -QSHPPSCPKDPCPNCSGCGWGGEECNCKTKIICAQCSHRRCGRSPDCCHNQCAAG 239
QY 244 CTGPKHSDCLACHFNHSGICELCHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNY 303
DB 240 CTGPRESDCLVQKQFDEATCKDTPPLMLNPTTYQMDVNPCKYSFGATCVKCKPRNY 299
QY 304 LSTDVGSCTLVCLPHNOVTAEDGTQRCCKSCPCARVCYGLGMOYIKANSKFTGITELE 363
DB 300 VVTDHGSVCVRACGPDYIEV-EEDGIRCKCKDGPCKVCNGIGIGIEFK-DTLSINATNIK 357
QY 364 -FAGCKKIFGSLAFPLPESFDGDPASNTAPLOPELOVFTLEETIGVLYISAWPDSLPLD 422
DB 358 HFYCTAISGDHLPLVAFKGSFTRTPPLDPRELEIKTKVKEITGELLQAWPDNWTDL 417
QY 423 SVFQNLQVIRGRILHNGAYSITLQGLGISWGLSLRSLRELASGLALIHNTHLCFVHTVPW 482
DB 418 HAPENLEIIRGTRKQHGFSVAVVGLNITSLGLSLKEISDGDVIIISGNRLCYANTINW 477
QY 483 DQLFRQVIKANSKFIGITELECEVGEGLACHOLCARGHCWGPPTQCVNCQSOFLRGQSCVE 542
DB 478 KCLFGTNPQTKIMNNRAEKDKAVNVHVCNPLCSSEGCWGPEDRCVSCQNVSRGECVE 537
QY 543 ECRVLQGLPREYVNAHCLPCHPCQPNQSVTCFGEADQCVACAHYKDPFPFCVARCPS 602
DB 538 KCVILEGEPRFVENSECICHPECLPQAMNITCTGRGPNQICQAHYIDGPHCVKTCPPA 597
QY 603 GVXPDLISYMPIKFPDEBEGACQPCPINCTHSCVDLDDKGPAPAEQASPLTSIVSAVVGIL 662
DB 598 GINGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCVEWPSGPKPSIATGIUGGL 656
QY 663 LVVVLGVVFGI-LIKRQOQKIRKVTMRLLQETELVEPLTFSGAMPNQAOIRLKETELR 721
DB 657 LFIWV-VALGIGLGFMRERHVRKTRLLQERLEVEPLTPSGEAPNQAHRLILKETEFK 715
QY 722 KVKVLGSGAGTGVYKGIWI PDGENVKIPVAIKVIRENTSPKANKEILDEAYVAVGVSPI 781

```

```

DB 716 KIKVLGSGAGTGVYKGIWIPEGEKVIPVAIKELREATSPKANKEILDEAYVAVSDNPH 775
QY 782 VSRLLGICLSTVOLVTQLMPEYGLLDHVRNRRGLSGQDLLNWCQIAKAGMSYLEDVRL 841
DB 776 VCRLLGICLSTVOLITQLMPEYGLLDYVRHKONIGSQYLLNWCQIAKAGMYLEDRL 835
QY 842 VHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRFTH 901
DB 836 VHRDLAARNVLKTPQHVKITDFGLAKLGAEEYHAEKGVPIKMMALESILHRIYTH 895
QY 902 QSDVMSYGVTVWELMTGAKPYDCIPAREIPDLLEKGERLPOPPICITDVYIMVYKCMWI 961
DB 896 QSDVMSYGVTVWELMTGSKPYDGIPASDISIILEKGERLPOPPICITDVYIMVYKCMWI 955
QY 962 DSECRPRFRELVSFSEMRARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLDDMDGLVDA 1020
DB 956 DADSRPKFERILIEFSKWARDPQRYLVIOGDERMHLPSPTDSNFYRALMDEDMEDVDA 1015
QY 1021 EYLVLPQOGFFCPDPAFGAGGMVHRHRSSTRSGGGDLTLGLEPSEEEA PRSPAPSEG 1080
DB 1016 DEYLIPOOGFF-----NSPST-----SRTPLLSLS 1041
QY 1081 AGSDVFDGDLGMGAAGKGLQSLPTHDPSPLOQRYSEDPTVPLPSET--DCYVAPLTCSPQPE 1138
DB 1042 ATSN-----NSTVACINRNGSCRVKEDAFLOQRYSSDPTGAVTEDNIDDAFL-----PVPE 1091
QY 1139 YVNOPDVRPPSPREGPLPAARPAAGATLERAKTSLPGKNGVWDXVAFAGGAVENPEYL- 1197
DB 1092 YVNO-SVPKPPAGSVQNPVHNQPLHP-----AFGRDLHYQN--PHSNAVGNPEYLN 1140
QY 1198 TPQCGAAPPHPPPAFSPAFDNLYWDO-----DP-----PERGAPPSTFKGTP 1241
DB 1141 TAQ-----PTCLSGGNSPALVIQKSHQMSLDNPDYQDFFPKETKENGIFKG-P 1190
QY 1242 TAENPEYGLDVP 1254
DB 1191 TAENAEYLRVAPP 1203

RESULT 6
ERB4_HUMAN
ID ERB4_HUMAN STANDARD; PRT; 1308 AA.
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (tyrosine kinase-type cell surface receptor HER4).
DE ERBB4 OR HER4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Foy L., Neubauer M.G., Shoyab M.;
RA "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT epidermal growth factor receptor family.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
RC TISSUE=Fetal brain;
RX MEDLINE=97476287; PubMed=9334263;
RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
RA Klagsbrun M.;
RA "A novel juxtamembrane domain isoform of HER4/Erbb4. Isoform-specific
RT tissue distribution and differential processing in response to
RT phorbol ester.";
RL J. Biol. Chem. 272:26761-26768(1997).

```



```
FT DOMAIN 186 334
FT DOMAIN 496 633
FT DOMAIN 718 985
FT NP_BIND 724 732
FT BINDING 751 751
FT ACT_SITE 843 843
FT DISULFID 189 197
FT DISULFID 205 205
FT DISULFID 213 221
FT DISULFID 217 229
FT DISULFID 230 238
FT DISULFID 234 246
FT DISULFID 249 258
FT DISULFID 262 289
FT DISULFID 293 304
FT DISULFID 308 323
FT DISULFID 326 330
FT DISULFID 503 512
FT DISULFID 507 520
FT DISULFID 523 532
FT DISULFID 536 552
FT DISULFID 555 569
FT DISULFID 589 577
FT DISULFID 580 589
FT DISULFID 593 614
FT DISULFID 617 625
FT DISULFID 621 633
FT MOD_RES 1162 1162
FT MOD_RES 1188 1188
FT MOD_RES 1258 1258
FT MOD_RES 1284 1284
FT CARBOHYD 138 138
FT CARBOHYD 174 174
FT CARBOHYD 253 253
FT CARBOHYD 358 358
FT CARBOHYD 410 410
FT CARBOHYD 473 473
FT CARBOHYD 495 495
FT CARBOHYD 548 548
FT CARBOHYD 576 576
FT CARBOHYD 602 602
FT CARBOHYD 1062 1062
FT CONFLICT 1080 1082
FT CONFLICT 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;

Query Match
Best Local Similarity 44.9%; Pred No. 9.9e-149; Length 1308;
Matches 607; Conservative 189; Mismatches 386; Indels 170; Gaps 30;

QY 1 MELA-ALCRWGLLL--ALLPFGAASQVCTGTDMKRLPASPETHLDMRLHLYQGVVQ 57
DB 1 MKLATGLWMGSLVAARTVQPSASQSCAGTENKLSLSLDLEQQYRALRKYENCEVVM 60
QY 58 GNLETLPTWASLFQDIEQVQYVLIANQVQVPLQRLIRVGTQOLFEDNVALVL 117
DB 61 GNLEITSIEHRDLSFLRSIREVTGYLVALNQFRLPLENLRIRGTKLVEDYALAI 120
QY 118 DNGDPLNNTPTVTGASPGRLQLRLSLTEILKGVLIQVLPOLCYQDTILWKDIFHKN 177
DB 121 LNYRKGNF-----GLQELGNLTELINGVYVDQNKFLCYADTIHQDILVRNP 171
QY 178 QLALTIDNTRSRACHPCSKGRGWGESSEDCQSLTRTVACGGC-ARCKGPIPTDC 236
DB 172 PSNMTLVSTIGSGGCRCHKSCG-RCWGPTENHCQTLTRTVACBQCDGRCYGPVSDCC 230
QY 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGYTPGASC 296
DB 231 HRECAAGCGSKDTCDFACMFNDSGACVTCQPTFVFNPTTFLQEHFNKATYTGAF 290
QY 297 TACPYNILSTDVGSCTLLVCLPHNQVETADBDGTQRCCKSKPCARVCYGLGHQYIKANS 356
DB 291 KKCPHNFV-VDSSSCVRACPSKKEV-EENGIKWCKPCTDICPKACDGI GTGSLMSAQT 348
```

```
QY 357 IGITELEPAGCKKIFGSLAFLPESPDGPASNTAPLQPELOQVFTLEBITGLYISAMP 416
DB 349 DSSNIDKINCTKINGNLIFLVGTGHGYPYNAIDDAIDPEKLVNFRVREITGFLNIQTWP 408
QY 417 DSLPDLSPQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLSRLSGSLALIHNNHLCF 476
DB 409 PNMTDFSFSNLVTIGRVLVSLGLLLIKQOGITSLQFQSLKEISAGNIYITDNSLVCY 468
QY 477 VHTVPMDOFL-----ROYIKANSKFIGITELECEVGEGLACHOLCARGHCWGPQTQVNC 531
DB 469 YHTINWTTLFTVNQIRIVIRNRR-----AENCTAEGMVNHLCSNDGCGPQDCLSC 523
QY 532 SQFLRGQCEVBEVRVLQGLPREYVNAHCLPCHPECQ-ONGSVTCFGEPAQOCVACAHY 590
DB 524 RFSRGKICIESCNLYDGEFREFENGSCVCECDSCQEKMEDGLLTCHGPGPNDCTKCSHF 593
QY 591 KDPPECVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGC----- 642
DB 584 KDGPNCEKCPDLVQGANSF--IFKYADQDRECHPCPNCTQCCNGPTSHDCIYYPTWGH 641
QY 643 ---PAQRASPLTSTVSADV-GILLVVLGVVFGILIKRQOKIRKYMRLLOETELVE 698
DB 642 STLPOHAR-TPL--IAAGVIGLFLVIMALTFAVYVRRKSTK-KKRALRRFL-ETELVE 696
QY 699 PLTPSGAMPNOAQMRLKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLREN 758
DB 697 PLTPSGTAPNAQRLKETELRKVKVLGSGAFGVYKGIWIPDGETVKIPVAIKILNET 756
QY 759 TSPKANKEILDEAYVMAGVSPYSRLLGICLTSTVQLTQMLPYGCLLDHVRNRRGLG 818
DB 757 TGPKANVEFMDEALITMASVDHPLVRLGLVCLSPITQLVTQMLPHGCLLEYVHEHKDNIG 816
QY 819 SODLLNWCQJAKGMSYLEDVRLVHRDLAARNLVKSPNNHVKITDGLARLLDIDETEVH 878
DB 817 SOLLNWCQJAKGMSYLEDVRLVHRDLAARNLVKSPNNHVKITDGLARLLDIDETEVH 876
QY 879 ADGKVPILKMALESILRRRTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEK 938
DB 877 ADGKVPILKMALESILRRRTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEK 936
QY 939 ERLPQPPCTIDVIMVVKCMWIDSECRPRFELVSEFSEMRARDPQRFVIONED-LGPA 997
DB 937 ERLPQPPCTIDVIMVVKCMWIDSECRPRFELVSEFSEMRARDPQRFVIONED-LGPA 996
QY 998 SPLDSTEYRSLLEDDMDGLVDAEYLVQCGFFCPDP----- 1035
DB 997 SPNDSKFFQNLDBEDLEDDMDAEYLVQ-QAFNIPPIY-SRTRIDSNRSRIGHSPPPA 1055
QY 1036 -APGAGGMVHRHRSSTRSGGGLTLGLPSEEAAPRSLAPSEAGSDVDGDLGMA 1094
DB 1056 YTPMSGSQFVYQDGFATQGG---MPMPYATTSTIPEAPVA--OGATAEMFDDSCNGT 1110
QY 1095 AKGLQSLPHTDPSPLQRYSEDTVPPLPS-----ETDGYVAPLTCSPQEVYNQDVRP 1147
DB 1111 LRKPVVHVQSDSSQRYSDPTVPAPERNPRAELDEEGYMTPMHDKPQEVYNPVE--- 1167
QY 1148 QPPSPREGPLAARPAAGATLERAKTSLSPKNGVVKDVFAGCGAVENPEVLTQGGNAPO 1207
DB 1168 -----ENPFVSR-----KNGDLQ-----ALDNPEYHSASSG----- 1194
QY 1208 HPPPA-----FSPAFONLYYWDQDPPERGA- 1232
DB 1195 -PPRADEYVNEPLYLNTFTNALGNAEYMKNSLLSVPEKAKAFDNPYWHLSLPRSTL 1253
QY 1233 -PPSTFKGTPT-----AENPEYL 1249
DB 1254 QHPDYLQESTYKFKYKQNGIRIRPIVAENPEYL 1285
```

RESULT 8
XMRK_XIPMA
ID XMRK_XIPMA
AC P13388;
STANDARD; PRT; 1167 AA.

DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
 GN XMRK OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OC NCBI_TaxID=8083;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=90015140; PubMed=2797166;
 RX Wittbrodt J., Adam D., Malitschek B., Mauele W., Raulf F.,
 RA Telling A., Robertson S.M., Schartl M.;
 RA "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RT inducing Tu locus in Xiphophorus.";
 RL Nature 341:415-421(1989).
 RN [2]
 RP REVISION TO 515.
 RA Schartl M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X16891; CAA34770.2; -
 DR PIR; S06142; S06142.
 DR HSP; P11362; 1FGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00220; S_TK; 1.
 DR SMART; SM00219; TYRK; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167
 FT MELANOMA RECEPTOR PROTEIN-TYROSINE
 FT KINASE.
 FT DOMAIN 26 642
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 643 665
 FT POTENTIAL.
 FT DOMAIN 666 1167
 FT CYTOPLASMIC (POTENTIAL).
 FT PROTEIN_KINASE
 FT NP_BIND 716 724
 FT ATP (BY SIMILARITY).
 FT BINDING 743 743
 FT ATP (BY SIMILARITY).
 FT ACT_SITE 835 835
 FT BY SIMILARITY.
 FT DISULFID 195 204
 FT BY SIMILARITY.
 FT DISULFID 199 212
 FT BY SIMILARITY.
 FT DISULFID 220 228
 FT BY SIMILARITY.
 FT DISULFID 224 236
 FT BY SIMILARITY.

FT DISULFID 237 245
 FT BY SIMILARITY.
 FT DISULFID 241 253
 FT BY SIMILARITY.
 FT DISULFID 256 265
 FT BY SIMILARITY.
 FT DISULFID 269 296
 FT BY SIMILARITY.
 FT DISULFID 300 311
 FT BY SIMILARITY.
 FT DISULFID 315 330
 FT BY SIMILARITY.
 FT DISULFID 333 337
 FT BY SIMILARITY.
 FT DISULFID 504 513
 FT BY SIMILARITY.
 FT DISULFID 508 521
 FT BY SIMILARITY.
 FT DISULFID 524 533
 FT BY SIMILARITY.
 FT DISULFID 537 553
 FT BY SIMILARITY.
 FT DISULFID 556 569
 FT BY SIMILARITY.
 FT DISULFID 560 577
 FT BY SIMILARITY.
 FT DISULFID 593 615
 FT BY SIMILARITY.
 FT DISULFID 618 626
 FT BY SIMILARITY.
 FT CARBOHYD 622 634
 FT BY SIMILARITY.
 FT CARBOHYD 114 114
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 144 144
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 201 201
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 356 356
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 365 365
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 398 398
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 621 621
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;
 Query Match 39.3%; Score 2674.5; DB 1; Length 1167;
 Best Local Similarity 45.1%; Pred. No. 1.3e-135;
 Matches 572; Conservative 166; Mismatches 391; Indels 139; Gaps 29;
 QY 4 AALCRWGLLLALLPPCAAST----QVCTGDMKRLRPASPETHLDMLRLHYQCGVQGN 59
 DB 8 AALLQ--LLLVLSISCCSTDPKVCQGTSMQTM---LDNHYLKMKNYSGCNVVLEN 62
 QY 60 LEITYLPTNASLFLQDIOEVQGVYLIAHNQVRQVPLQRLRIVRGTQOLFEDNVALAVLDN 119
 DB 63 LEITYTQENQDLSFLQSIQEVGVYLIANNEVSTIPLVNLRLIRGQNLVYEGNFTLLVMSN 122
 QY 120 GDLNNTPTVTGASPGGLRELQLRSITELKGVLIORNPQLCYQDTILWKDIFHKNNQL 179
 DB 123 YOK-NPSSP--DVYQVGLKQLQLSNLITELSGVKVSHNPLLCNVETINWMDIVDKTNP 179
 QY 180 ALTLDTNESRACHPCSPMKSGRCSGSESEDCOSLTRVCAGGC-ARCKGPLPTDCHE 238
 DB 180 TWNLIPIHAFERQCKCDHCGVNGSCWAPGFGHCQKFTKLLCABQCRRRCRGPIDCCNE 239
 QY 239 QCAAGCTGPKHSDDLACLFHNSHIGICELHCPALVYNTDTTFESMPNPEGRYTFGASCVTA 298
 DB 240 HCAGGCTGPRATDCLACRDFNDGDKTCTPPKIYDIVSHQVVDNPNIKYTFGAACVKE 299
 QY 299 CPYNILSTDVSGCTLVCPILHNOEVTAEDGTQRCCKSPCARVCYGLGMQYIKANSKFIG 358
 DB 300 CPSNYVWTE-GACVRSKCSAGMLEVD-ENGKRSCKPCDGVCPKVDGIGISL-SNTIAVN 356
 QY 359 ITEL-EFACKKIFGSLAPLPEFSGDPRASNTAPLOEQLOVFETLEETGYLIYISAWPD 417
 DB 357 STNIRSFNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLNLTITTYKEITGYLIVMWPE 416
 QY 418 SLPDLISVFONLQVIRGRILHNGAYS-LTLOGLGISWLGRLSLRSLGSLALIHNTLCLF 476
 DB 417 NMTLSLVFONLEIRGRITFSRGFVQVVRHLQWLGURSLKEVSGAGNVILKNTQLRY 476
 QY 477 VHTVPWDLFRQYIKANSKFIGITELECYEGGLACHQLCARGHGWGPGTQCVNCSQFLR 536
 DB 477 ANTINWRRLFRSEDQS-----IEVDARTENQTCNNECEDGCGPGTMCVSLHVDR 529
 QY 537 GQECVEECRVLOGLPREYNARHCLPCHPECPQNGSVTCFQPEADQCACAHYKDPDFC 596
 DB 530 GGRCVASCNLLQGEPRQAQVDRGRCVQCHQECVQLVQDLSLTCTCYGPGPANCSSKAHFQDQPC 589
 QY 597 VARCPGKVPDLSPYMPIWKFDPDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIYS 656

```
Db 590 IPRCPHGILGDDTL-IWKYADKMGQCQCHONCTQCGSGPLSGCRGD-IVSHSLAVG 647
Qy 657 AVGILLVVLGVFGILIKRRQKIRKYTHRRLLQETELVEPLTPSGMNPQOMRLK 716
Db 648 LVSGLLITVIVALLIVLLRRRIK-RKTRICLLQEKELVEPLTPSGAENQAFRLK 706
Qy 717 ETELKRVKLGSGAGFYVYGIWIPDGENVPIPAIKVLRENTSPKANKETLDEAYVAG 776
Db 707 ETEFKDORVLGSGAGFYVYKGLWNPGENIRIPAIKVLRENTSPKANKETLDEAYVAG 766
Qy 777 VGSYVSRLLGICLTSTVQLVTQMPYCGCLLDHVRNRLGSGODLNCWQIAKMSYL 836
Db 767 VDHPVCRLLGICLTSAVLVTQMPYCGCLLDYVRQHQERICGQWLLNCWQIAKMSYL 826
Qy 837 EDVPLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKMALESILR 896
Db 827 EERHLVHRDLAARNVLLKNPNHVKITDGLSKLLTADEKEYQADGGKVPKMALESILQ 886
Qy 897 RRFTHQSDVMSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDYVIMV 956
Db 887 WTYTHQSDVMSYGVTVVWELMTFGSKPYDGIIPAKIASVLENGERLPPOPICTIEVYMI 946
Qy 957 KCMWIDSECRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMGD 1016
Db 947 KCMWIDPSSRRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMGD 1001
Qy 1017 LVDAEYVLPQGGFCFDPAPGAGGWHHRHSSTRSGGGLTLGLPSEEAAPRPLA 1076
Db 1002 VVDAEYVLPYKRI-----NRQGS-----EPC 1024
Qy 1077 PSEGAGSDVDFDGLGMAKGLQLTHDPSPLQRYSEDPTV-PLPSETDGYVAPLTCSP 1135
Db 1025 PPTGH-----PVRENSITLRNISDPTQNALEKLDGH----- 1056
Qy 1136 QPEYVQDVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTUSPKNGVVK 1182
Db 1057 --EYVNPFGSTSSRLSDIYNPNYEDLTGNGVPSLSQEAETNFSRPEYLTNQNSL-- 1112
Qy 1183 DVFAFGAVENPEYLTQGGAAPOHPHPPAFSPAFDNLVYWDQPPPERGAPPSFTKGTPT 1242
Db 1113 -PLVSSGSMDDPY---QAG-----YQAAF-----LPQTGAULTGNGMFLPA 1149
Qy 1243 AENPEYLG 1250
Db 1150 AENLEYLG 1157

RESULT 9
ERB3 HUMAN
ID -ERB3 HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 19, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
```

```
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RC MEDLINE=93282822; PubMed=7685162;
RX Katoh M., Yazaki Y., Sugimura T., Terada M.;
RA "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RT Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTA.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M34309; AAA35979.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSSP; P11362; 1FGK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN KINASE TYR; FALSE_NEG.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
DR Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
DR Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202
```



```

FT DISULFID 210 BY SIMILARITY.
FT DISULFID 214 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 141 183 EILSGGVYIEKDKLCHMDTIDVRDRAEIVVKNDR
SC -> GQFNPVSPGLPQPAQDWYLLDDDDPRLLTLLSSSK
VPVTLAAV (IN SHORT ISOFORM).
MISSING (IN SHORT ISOFORM).
E -> G (IN REF. 2).
E -> G (IN REF. 2).
SEQUENCE 1342 AA; 148097 MW; 7201ETP66CA374BD CRC64;

```

Query Match 35.2%; Score 2393.5; DB 1; Length 1342;
 Best Local Similarity 40.0%; Pred. No. 1.5e-120;
 Matches 530; Conservative 194; Mismatches 445; Indels 155; Gaps 35;

```

Qy 10 GLLLALLPPGNA--STOVCTGDMKRLRASPETHLMRLHYOGQVQGNLELTLYPT 67
Db 11 GLVFLARGSEVGNQAVCPGLNGLSVTGDENQYQTLKLYERCEVMGNLEIVLTGH 70
Qy 68 NASLSFLQDIQEVGYVLIAHNQVQVPLQRLRIVRGTLQFEDNYALAVLDNGPLANNNT 127
Db 71 NADLSFLQWIREVTGYVLVAMNEFTLPLNLRVVRGTQYVDGKFAIFVM-----LNYNT 125
Qy 128 PVTGASPGGLRELQRLSLTEILKGVLIQRNPOLCYODTILWKDIFHKNNQLALTLDTN 187
Db 126 ----NSSHALRQLRLTQLTEILSGVYIEKNDKLCHMDTIDWRDIVDRD---AEIVVKD 178
Qy 188 RSRACHPCSPMKGSRGWCSSGESSDCSLTRTVACAGC-ARCKGPLPTDCHEQCAAGCTG 246
Db 179 NGRSCPPCHEVCKG-RCWGPGSEDCQLTKTICAPQCNHGCHFGPNPQCCHDECAGCGS 237
Qy 247 PKHSDCIACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNYVLT 306
Db 238 PQDTDCFACHFNDGACVRCPOPLVYNKLTFOLEPNHTKYQYGVGVVASCCHPNFV-V 296
Qy 307 DVGSCTLVCPHMQEVAEDGTORCEKSKPCARVCYGLGMQYIKANSKF--IGITELE- 363
Db 297 DOTSCVRACPPDKMEVD-KNGLKXCEPCGGLCPKACEGTG-----SGSRFQTVDSNIDG 350
Qy 364 FACKKIFGSLALFPESFDCDPSANTAPLOPEQLOVFETLEETGYLYISAWPDSLPDLS 423
Db 351 FVNCTKILGNLDELITGLNGDPMWHKIPALDPEKLNVRFTREITGYLYIQSWPPHMFNS 410
Qy 424 VFQNLQVIRGRILHNGAYS-LTQGLGISWGLRSLRGLSGLALIHNTLHCFVHTVPV 482
Db 411 VFSNLTIGRSYLNRCGFSLLIKNVLNVTSLGFRSLKEISAGRIYISANRQLCYHSLNW 470

```

```

Qy 483 DQLFRQYIKANSKFIGITE-----LECVGEGIACHQLCARGHGWGPGTQCVMC 531
Db 471 TKVLR-----GPTTEERLDIKHNRPRRDCVAEGKVCDCPLCSSGGCWGPGQCLSC 520
Qy 532 SQFLRGQECVECRVLQGLPREYVNAHCLCHPECCQNGSVTCTGFEADOCVACAHYK 591
Db 521 RNSRGGVCTHNCNLFNGEPREFAHEACFSCHEPCQMEGTATCGSGSDTCAQCAHFR 580
Qy 592 DPPFCVARCPSGVKPDLSYMPIWKFPDBEAGACQPCPINCTHSCVDLDDKGCPEAQR- 648
Db 581 DGHCVSSCPHGVLG--AKGPIYKYPDVQNECRCHENCTQCGKGPDLQDGLQTLVLIG 638
Qy 649 -SPLTSIVSAVVGILLVVVLGVFGIILKRRQOKIR-KYTMRRLLQETELVEPLTPSGAM 706
Db 639 KTHLTMAITVIAG--LVVIFMLGCTFLYWRGRIQNKRRMRYLERGESIEPLDPS-EK 695
Qy 707 PNOAQMRLIKETELRKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKE 766
Db 696 ANKVLARIFKETELRKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKE 766
Qy 767 ILDEAYVMAGVSPVSRLLGICLTSTVQLVTQLMPIYGLCLLDHVRENRRGLSGDQLLNC 826
Db 756 VTDEMLAIGSLDHAHIVRLGLCPGSSQLQVTLPLGSLLDHVRRHGRGALGPQLLWNG 815
Qy 827 MQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPI 886
Db 816 VOIAKGMVYLEEHGMVHRNLAARNVLKSPQVQVADGVADLPPDDKQLLYSEAKTPI 875
Qy 887 KMALESILARRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPI 946
Db 876 KMALESIFHGKYTHQSDVMSYGVTVWELMTFGAEPYAGLRLAEVDPDLEKGERLAQPI 935
Qy 947 CTIDVYMWKCMIDSECRPRPRELVESESRMARDPQFVVIQNEIDLGA---SPLDST 1003
Db 936 CTIDVYMWKCMIDENIRPTFKELANEFTRMARDPPRYLVIKRES-GPGIAPGPEPHG 994
Qy 1004 FYRSLLEDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGL 1063
Db 995 LTNKKEVELEPELDDLDLEAED-----NLATTILGSALSLPV 1035
Qy 1064 EP-SEBAPRSLAPSEGAGSDVFDGLGMAAGKLOSLPHD-PSPLQRYSESDTVPVLP 1121
Db 1036 GTLNRPRGSOALLSPSSGY-MPMNQNLGSECSAVSGSSERCPRPVSLH-----BMP 1088
Qy 1122 -----SETGYVA-----PLTCSPOPE-----YUNQPDVPRPQPSREG 1155
Db 1089 RGCLASSESSEGHVTGSEAELOEKVSMCRSRSRSPRGRDSAYHSQRHSLLTPTVPLSP 1148
Qy 1156 P-----LPAARPAGATLERAKTLSP-GKNGVV-----KDVFAFGAVENPEYL 1197
Db 1149 PGLEEDVNGVMPDTHLKGTPSSREGTSSVGLSSVLGTDEED-----EYEYM 1200
Qy 1198 TPQGGAAPOPHPPAFSPAFDNLVYWD-----QDPPERGAPPSTFKGPTTAEN 1245
Db 1201 NRRRRHSP-PHPPRPSLEELGYEYMDVGSLSASLSTQSCPLHPVIMPTAGTTDDED 1259
Qy 1246 PEYL 1249
Db 1260 YEYM 1263

```

RESULT 10
 ERB3 RAT

ID "ERB3_RAT STANDARD; PRT; 1339 AA.
 AC Q62799; Q62955;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
 DE (c-erbB3).
 OS ERBB3.
 GN Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```

Db      520 CRNYSREGVCVTHCNFQGEPRFVHEAQCFSCHEPLMEGTSYTCNGSGSDACARCAHF 579
Qy      591 KDPFFCVARCPGSGVKPDLSTYMPINWKFDPDEGACOPCPINCTHSC--VDLDDKGPAPORA 648
Db      580 RDGPHCVNSCPHGILG--AKGPIYKYPDAQNECPCHENCTQCGNFELQDCLGOAEVLM 637
Qy      649 SPLTISVAVGILLVVLGVVFGILKRRQOKIR-KYTWRRLLQETLEVEPLTPSGAMP 707
Db      638 SKPHLVIAVTVG--LAVILMLGSGFLYWRGRRITQNKRAMRRYLERGESIEPLDPS-EKA 694
Qy      708 NOAQMRILKETELRKVKVLGSGAGFTYVKGWIIPDGENVKIPVAIKVIRENTSPKANKEI 767
Db      695 NKVLARIFKETELRKVLGSGVFGTVHKGWIPEGESIKIPVCIKVIEDKSGRQSFQAV 754
Qy      768 LDEAYNVAGVSPVYRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGLSGQDLNWM 827
Db      755 TDHMLAVGSLDHAHIVRLGLCPGSLQLVTOYPLGLSLLDHVKQHRETILGPQLLNWGV 814
Qy      828 QIAKGSYLEDLVRLVHDLAARNVLKSPHVKITDFGLARLLDIDETEHADGKVPPIK 887
Db      815 QIAKGMVYLEHSHVHRLDLARNMLKSPSQVQVADFLPPDKQLLHSEAKTPIK 874
Qy      888 WMALESLRRRTHQSDVWSVGVTVWELMTFGAKPYDGPARETPDLLEKGERLPQPPIC 947
Db      875 WMALESHFGKYTHQSDVWSVGVTVWELMTFGAEPYAGLRLAETPDLLEKGERLAQPOIC 934
Qy      948 TIDVYIMVWKWIDSCRRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRS 1007
Db      935 TIDVYVWKWCMIDENIRTPFKELANEFTMRADPPRYLVIKRAS-GPGCTP--PAEAPS 991
Qy      1008 LLEDDMGDLVDAEYLVPOGFPFCDPAPGAGGVMHRRSSSTRSGGGDLTLGLRPS 1067
Db      992 VLTTKEL-----QEALELEP-----AGSDVFDGDLGCAAGLQ 1017
Qy      1068 E-----EAPRSLAPSEG-----AGSDVFDGDLGCAAGLQ 1099
Db      1018 EGLATSLGSLALPTGLTLPRGSQSLSSFGVMPNQSLGACLDLSAVLGRQFSR 1077
Qy      1100 SLPHDPSPLQRYSEDTPVLPSTDGCV---APL-----TC-----SPOPE---Y 1139
Db      1078 PISLH-PIPRGR-----PASESGHVTGSEALQEKVSVCKRSRSPRGRDSAY 1129
Qy      1140 VNPQDVRPQPPSPREGP-----LPAARPAGATLERAKTLSP-GKNGV----- 1181
Db      1130 HSQRHSLTLTPTVPLSPGLEEDGNGYVMPDTHLRGASSRREGTLSSVGLSLVLTREED 1189
Qy      1182 KDVFAGCGAVENPEYLTPOGGAAPQPHPP 1210
Db      1190 ED-----EYEVYMKRKGSP-PRPP 1209

RESULT 11
EGFR_DROME STANDARD; PRT; 1426 AA.
AC P04412; O61501; Q9W2G0; P91868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schupbach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
that several genetically defined classes of alleles cluster in

```

```

RT subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
[2]
RN REVISIONS.
RP Clifford R., Schupbach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
[4]
RN SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RP STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
[5]
RN SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
[6]
RN SEQUENCE FROM N.A. (ISOFORM TYPE I).
RP STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Anatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Arkil J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dinkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Waberman D.A., Weinstein G.M., Weissenbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[7]
RN SEQUENCE OF 959-1078 FROM N.A.
RP

```

RC STRAIN=Daekwanryeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal
RT growth factor receptor.";
RL Nature 314:178-180(1985).
RN [8]
RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP ANALYSIS.
RX MEDLINE=92038942; PubMed=1936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RT "interallelic complementation among DER/flb alleles: implications for
RT the mechanism of signal transduction by receptor-tyrosine kinases.";
RL Genetics 129:191-201(1991).
RN [9]
RP REVIEW.
RX MEDLINE=97248481; PubMed=9094709;
RA Perrimon N., Perkins L.A.;
RT "There must be 50 ways to rule the signal: the case of the Drosophila
RT EGF receptor.";
RL Cell 89:13-16(1997).
CC -!- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS.
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES. MAINTENANCE OF AMNIOSEROSA
CC AND VENTRAL NEUROECTODERMAL CELLS. GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICULE.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF052754; AAC08536.1; -;
DR EMBL; AF052753; AAC08536.1; JOINED.
DR EMBL; AF052754; AAC08535.1; -;
DR EMBL; AF052752; AAC08535.1; JOINED.
DR EMBL; K03054; AAA51462.1; -;
DR EMBL; K03417; AAA51460.1; -;
DR EMBL; K03416; AAA50965.1; -;
DR EMBL; K03418; AAA51461.1; -;
DR EMBL; AF109077; AAD26134.1; -;
DR EMBL; AF109078; AAD26132.1; JOINED.
DR EMBL; AF109082; AAD26132.1; JOINED.
DR EMBL; AF109084; AAD26133.1; -;
DR EMBL; AF109087; AAD26133.1; JOINED.
DR EMBL; AF109089; AAD26133.1; -;
DR EMBL; AF109079; AAD26130.1; -;
DR EMBL; AF109081; AAD26130.1; JOINED.
DR EMBL; AF109079; AAD26131.1; -;
DR EMBL; AF109083; AAD26131.1; JOINED.
DR EMBL; AF109080; AAD26131.1; JOINED.
DR EMBL; AF109080; AAD26135.1; -;

DR EMBL; AE003454; AAF46732.1; -;
DR EMBL; X02293; CAA26157.1; -;
DR EMBL; X78920; CAA55523.1; -;
DR EMBL; X78918; CAA55521.1; -;
DR EMBL; X78919; CAA55522.1; -;
DR PIR; A00640; GQFFE.
DR HSSP; P11362; IFGK.
DR FlyBase; FBgn0003731; Egfr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW Developmental protein.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 869 889 POTENTIAL.
FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 938 1198 PROTEIN KINASE.
FT NP_BIND 944 952 ATP (BY SIMILARITY).
FT BINDING 971 971 ATP (BY SIMILARITY).
FT ACT_SITE 1063 1063 BY SIMILARITY.
FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
Query Match 28.3%; Score 1925; DB 1; Length 1426;
Best Local Similarity 32.5%; Pred. No. 1.7e-95;
Matches 466; Conservative 187; Mismatches 422; Indels 360; Gaps 45;
QY 24 QVCTGTDMKRLPASPTLHDLMLHLYOGQCVQVGNLELTYPPT-NASLSFLQDIOEVQOG 82
DB 100 KICIGTKSRLSPVSNKEHHYRNLRDRYTNCTYVDGNLKLTLWLPNENLDSLFDNIREVTG 159
QY 83 YVLTAHQVROVPLQRLRIVRGTOF-----EDNYALAVLNGDPLNNTTPVTGASPGGL 137
DB 160 YILISHVDVKKVFPKLIQINGRTLFLSVSEEEKYALFV-----TYSKM 203
QY 138 RELQLRSITELKGGVLIQRNPQLCYQDTILWKDIFHKNNQLALTLIDNRSRACHPCSP 197
DB 204 YTLLEIPDLRVNLGVQVGHNNYNLCHWRTIOWSIVNSGTDAYNYNDFATPERECPKCHE 263
QY 198 MCKSGRWGESSEDCQSILTRTVAGGCA--RCKGPLPTDCHEQCAACTGPKHSDCLAC 255
DB 264 SCTHG-CWGECPKNCQKFSKLTCSPOCAGGRCYGPKPKECCHLFCAGGCTPTQKDCIAC 322
QY 256 LHENHSGICELHCPALVTYNTDTESNPNEGRVTFGASCYVACPNYVLSLTVGSCITLVC 315
DB 323 KNFFDEAVSKEECPMKRYNPTTVLETNPEKGYAYGATCVKECP-GHLLRDNGACVRSC 381
QY 316 PLHNQEVTAEDGTQCEKCKSPCARVCYGLGMQVIKANSKFIGITEL-----EPAGCKK 369
DB 382 PQDKMDKGGE-----CVPNGCPKPKTC-----PGVTVLHAGNIDSRNCTV 422
QY 370 IFGSLAFLPESFDG--DPASNTA-----PLQPSQLQVFETLEBITGYLYYSAMPDSDLPD 421
DB 423 IDGNIRILDQTFSGFDQVYANYTMGRPIYIPLDPERREVFSTVKEITGYLVNIEGTHPQPN 482
QY 422 LSVFQNLQVIRGRILHNGAY-SLTLOGLGISWLGRLSRLGSLGLIHHNTHLCFVHTV 480
DB 483 LSYFRNLETIHGRLQSMESMFAALAIKVSLSYLEMRNLKQISSGVVIOHNRDLCCYVSN 542
QY 481 PWDQLFRQ-----YIKANSKFIGITELECEVGEGLACHQLCARGHCWGPGPTQCVCNCSQFL 535

```
Db 543 RWAIOKEPQKVMNENLR-----ADL-CEKNGTICSDQCNEGCMGAGTDQCCTCKNFN 597
Qy 536 RGCEVCEVRLQGLPREYVNAHCLPCHPECPONGSVTCFCGPEADQCACAHYKDPFP 595
Db 598 FNGTCTADCGYISWAYK--FDNRCKICHPECR-----TCNGAGADHCECVHVRDQGH 649
Qy 596 CVARCP-----SGVK----- 605
Db 650 CVSECPKNKYNDRGVCHRECHATCGCTGPKDTIGIGACTTCNLAIINNDATVXKCLLKD 709
Qy 606 --PDLNMPYWKPF--PDEEGACOP-----CPI-----NCTH----- 632
Db 710 KCFD-GY--FWEYVHPOEQSLKPLAGRAVCRKCHPLCELCCTNYGYHEQVCSKCTHYKRR 766
Qy 633 -----SC-----VDLDDKG----- 641
Db 767 EQCETECPADHYTDEORECFQHPCEGCTGPGADDDCKSRNFKLFANETGPNVNTM 826
Qy 642 -----CPAEOR-----ASPLTS-----IVSAVVGILLVVVLG 668
Db 827 FNCTSKCPLEMRHVNVOYTAIGPYCAASPPRSSKITANLDVNMFIITGAVLVPTCILC 886
Qy 669 VFGILIKRQOKIRKVT--MRRLQETELVEPLTPSGAMPNQAMQMRILKETELRKVKVL 726
Db 887 W--TYICRQKQAKKETVMWTWALSCEDESEPLRPSNIGANLCKLRIVKDAELRKGGLV 944
Qy 727 GSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYVMAGVSPYVSRLL 786
Db 945 GMAFGGVYKGVVPEGENVKIPVAIKELKSTGAESSEFLEAYIMASEEHVNLKLL 1004
Qy 787 GICLTSTVLQTLMPYGCILLDHVRENRLGSDLLNWCMIQAKGMSYLEDLVRLVHRDL 846
Db 1005 AVCSSQMLITQLMPLGCLLDVYNNRDKIGSKALLNWSQIAKGMSYLEKELVHRDL 1064
Qy 847 AARNVLKSPNHVKITDFGLARLLDDETEYHADGCKVPKWALESILRRRFTHQSDVM 906
Db 1065 AARNVLVQTPSLVKITDFGLAKLLSSDSNEYKAAGKMPKWLALCEIRNRVFTSKSDVM 1124
Qy 907 SYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVVMVMVKMWIDSECR 966
Db 1125 AFGVTWELLTFQRPHENIPAKDIPDLIEVGLKLEQPCSDIYCTLLSCWHLDAAMR 1184
Qy 967 PRFRELVSFMRWARPQRFVJONEDLG--PASPLDSTFYRSLLED--DMGDLVDAE 1021
Db 1185 PTEKQLTTVAEPARPGRYLAIPGDKFRLPA-----YTSODEKDLIRKLAPTTDGS 1237
Qy 1022 EYLVPOGFCPPAPCAGGVHRRHSSTRSGGDLTLGLFSEBEAP-----RSPL 1075
Db 1238 EAIKAPDDYLOPKAAPGPS-----HRTDCT-----DEMPKLNRYCKDPS 1276
Qy 1076 APSEGAGSDVFDG---DLGMAAGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLT 1132
Db 1277 NKNSSTGDDERDSAREVGVGNLR-----LDLPVDEDDYLMPT-T 1314
Qy 1133 CSPQPEVYNQDVRPOPPSPREGPLPAARPAAGATLERAKTLPSPKNGVWVDVFAFGAVE 1192
Db 1315 CQPGPNNNNMN-----NPNQNNMAVGAAGVM-----DLIGVPVSVD 1353
Qy 1193 NPEYL-----TPQGAAPQPH-----PPAPFSP-AFDNLYYWD 1224
Db 1354 NPEYLLNAQTGLGVGESPIPTQTIGIPVMGGPGTMEVKVPMPPGSEPTSSDHEYIND 1408
```

RESULT 12

```
ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
```

```
OS Avian leukosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
processing and promoter insertion result in expression of an
amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC
CC EMBL: M10066; AAA48763.1; ALT_INIT.
DR PIR: A00643; TVCHLV.
DR PIR: B00643; TVFVLV.
DR HSP: P11362; IFGK.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00219; TyKc; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
DR TRANSFERASE: PS00011; PROTEIN KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.7%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 1.6e-86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;

Qy 587 CAHYKDPPEFCVAPCGVGRPDLSYMPIWKPDEGACQPCINCTHSCVDLDDGCPAEQ 646
Db 3 CAHFIDGPHCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPNCTGCKGPGLEGCP--- 58
Qy 647 RASPLTSIVSAVV-GILLVVVLGVGVGILIKRQOKIRKVTMRLLQETELVEPLTPSGA 705
Db 59 NGSKTFSIAAGVVGGLLCLVVGGLGIGLYLRR-HIVKRTLRLQLQERLVEPLTPSGE 117
Qy 706 MPNQAMRILKETELRKVKVGLSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 765
Db 118 APNQAHRLILKETEFKVKVGLSGAFGTVYKGLWIPGEKVKIPVAIKELREATSPKANK 177
Qy 766 EILDEAYVMAGVSPYVSRLLGLCLSTVQLVTLQMLPYGCLLDHVRENRLGSLQDLLN 825
Db 178 EILDEAYVMASVDNPHVCRLLGLCLSTVQLITQLMPYGCILLDVIYREHKDNIGSQYL 237
Qy 826 CMQIAKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDDETEYHADGCKVP 885
```

```
Db 238 CVOIAKGNVLEERRLVHRDLAARNVLVKTQHVKITDGLAKLLGADEKEYHAEGGKVP 297
Qy 886 IKWMALESILRRRFTHQSDVMSGYVTWELMTFGAKPYDGIPIAREIPDLLEKGERLP 945
Db 298 IKWMALESILHRIYTHQSDVMSGYVTWELMTFGSKPYDGIPIAREIPDLLEKGERLP 357
Qy 946 ICTIDVYMWKMWIDSECRPRELSEFSESRMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYMWKMWIDSECRPRELSEFSESRMARDPQRFVVIQ-NEDLGPASPLDSTF 417
Qy 1005 YRSLEDDDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1065 PSEEAERSPV-----APSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVP 1119
Db 450 -----SRTPLLSLSLTSNNSATNCID-----RNGQGHVPVREDSFVQRYSDPTGN 495
Qy 1120 LPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPREGPLPAARPAAGATLERAKTSPGK 1177
Db 496 FLEESIDDDGFL-----PAPEYVQ--LMPKKPS-----TAMVQ 526
Qy 1178 NGVVKDVF-----AFGGAIVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVY 1222
Db 527 NQIYNNISLTAISKLPMDSRQYNSHSTAVDNPYL-----NTNOSPLAKTVFESSPY 578
Qy 1223 WDQ-----DPPE-----RGAPSTFGTPTAENPEYLGLDVP 1254
Db 579 WIOGSHQINLDNDPYQODFLPNETKPNGLLKVAENPEYLVRAAP 625
```

```
RESULT 13
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RN STRAIN=H;
RC SEQUENCE FROM N.A.
RX MEDLINE=84026539; PubMed=63113229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RL gene family.";
RL Cell 35:71-78(1983).
RN [2]
RN SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debierre B., Henry C., Benais M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
RL new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR PIR; K01216; AAA42400.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76EBCDD06745D609 CRC64;

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 4.7e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

Qy 587 CAHYKDPFFVCARCPGVKPDLSYMPIWKFDEGACQPCPINCSTHSCVDLDDKGPAAEQ 646
Db 3 CAHFIDGPHCVKACPAVGLGENDTL-VRYKADANAVCOLCHPCNCTRGCKGLEGCP--- 58

Qy 647 RASPLTSIVSAV--GILLVVVGVVFGILIKRQOKIRKYTMRLRLQSTELVEPLTPSGA 705
Db 59 NGSKTPTSAAGVGGLLCLVVGIGLYLRRR-HIVKRTLRLRLQSTELVEPLTPSGE 117

Qy 706 MPNQAQRIILKETELRKVVLGSGAFGVYKGIWIDGENVKIPIVAKVLENTSPKANK 765
Db 118 APNQHRLILKETEFKKVVLGSGAFGTIYKGLWIPGEKVKIPIVAKVLENTSPKANK 177

Qy 766 EILDEAYVMAGVSGPYVSRILGICLTSTVQLVTOLMPYGCLLDHVRENRLGSDLLNW 825
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVQLITOLMPYGCLLDYIREHKDNGISYLLNW 237

Qy 826 CMQIAKMSYLEDLVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGGKVP 885
Db 238 CVQIAKGMVLEERRLVHRDLAARNVLVKTQHVKITDGLAKLLGADEKEYHAEGGKVP 297

Qy 886 IKWMALESILRRRFTHQSDVMSGYVTWELMTFGAKPYDGIPIAREIPDLLEKGERLP 945
Db 298 IKWMALESILHRIYTHQSDVMSGYVTWELMTFGSKPYDGIPIAREIPDLLEKGERLP 357

Qy 946 ICTIDVYMWKMWIDSECRPRELSEFSESRMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYMWKMWIDSECRPRELSEFSESRMARDPQRFVVIQ-NEDLGPASPLDSTF 417

Qy 1005 YRSLEDDDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449

Qy 1065 PSEEAERSPV-----APSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVP 1119
Db 450 -----SRTPLLSLSLTSNNSATNCID-----RNGQGHVPVREDSFVQRYSDPTGN 495

Qy 1120 LPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPREGPLPAARPAAGATLERAKTSPGK 1177
Db 496 FLEESIDDDGFL-----PAPEYVQ--LMPKKPS----- 524

Qy 1178 NGVVKDVF-----AFGGAIVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVY 1222
Db 527 NQIYNNISLTAISKLPMDSRQYNSHSTAVDNPYL-----NTNOSPLAKTVFESSPY 1218
```



```

Db 525 --VQNIQYFISLTAISKLPWDSRYQNSHSTAVDNPEYL-----NTQSPKLTVP 574
QY 1219 NLYWDDPPPERGAPPSTFGTPTAENPEY 1248
Db 575 SSPYIQSGNHQ-----INLNDPDY 594

RESULT 14
ERBB_AVIEU
ID ERBB_AVIEU STANDARD; PRT; 540 AA.
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OS Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
ON NCBI_TaxID=103898;
RX SEQUENCE FROM N.A.
RX MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells."
RT Mol. Cell. Biol. 6:1751-1759 (1986).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M13179; AAA42401.1; -
CC PIR; A25231; TVFVEB.
CC HSSP; P11362; 1FGK.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00669; pkinase; 1.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00109; PROTEIN KINASE TVR; 1.
CC PROSITE; PS50011; PROTEIN KINASE DOM; 1.
CC Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
KW DOMAIN 132 399 PROTEIN KINASE.
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA068B65D CRC64;

Query Match 24.0%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Pred. No. 3.3e-80;
Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;

QY 587 CAHYKDPFVCARCPGVKPDLSYMPFIWPPDEGACPCPCINCTHSCVDLDDKGCPRQ 646
Db 3 CAHFIDGPHCVKACPAVGLGENDTL-VKMYADANAVCQLCHPCTRGCKGPGLEGCP--- 58
QY 647 RASPLTSIVSAVY-GTLLVVLGVFGILIKRQKIRKYTWRRLLQETELVEPLTPSGA 705
Db 59 NGSKTPTSIAAGVVGLLLVVGLGIGLYLRRR-HIVRKRTLRLRLQERLEVEPLTPSGE 117
QY 706 MPNQAMRILKETELRKVKVLGSGAFGTGKGIWIPDGENVKIPVAIKVLRNTSPKANK 765
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

```

```

Db 118 APQAHURLIKETEFKKVKVLGFGAFCTVTVKGLWIPGEKVTIPVAIKELREATSPKANK 177
QY 766 EILDEAYVMAGVGSPPYVSRLLIGLICLTSTVOLVLTQMPYGCGLLDHVRNRRGLSGQDLNWN 825
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 178 EILDEAYVMASVDNPHVCRLLIGLICLTSTVOLITQMPYGCGLLDYIREHKONIGSQYLLNW 237
QY 826 CMQIAKMSYLELDVRLVHRLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 885
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 238 CVQIAKGMNYLEERHVMVHRLAARNVLKTPQHVKITDFGLAKQLGADAEKEYHAEKGKVP 297
QY 886 IKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOP 945
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 298 IKWMALESILHRYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISVLEKGERLPPOP 357
QY 946 ICTIDVYIMVKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQ-NEDLGPASPLOSTF 1004
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 358 ICTIDVYIMVKCMWSDADSRPKFRELIAEFSKWARDPPRYLVITQGERMHLPSPTDSKF 417
QY 1005 YRSLLDDDDGMDLVDAEYLVPOQGFPCPDPAAGGAGGVHRRHRSSTRSGGDLTLGLE 1064
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1065 PSEEEAPRSL-----APSEGAGSDVFDGDLGMAAKGLQSLTHDPSLQRYSEDTVP 1119
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 450 -----SRTPLLSSLSATSNNSATNCIDRNG-----H----- 476
QY 1120 LPSETDGYVAPLTCSPQPEYVQPDVVRPQPPSPREGPLPAARPAGAT-LERAKTLPQKN 1178
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 477 -PVREGDL-----PAPEYVQ--LMPKFPSTAMVQNIYVYISLTAISKLPIDRSYQN 527
QY 1179 GVVKDVPFAGGAVENPEYL 1197
Db 528 -----SHSTAVDNPEYL 539

RESULT 15
EGFR_CHICK
ID EGFR_CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
DE (Fragment).
DE EGFR.
OS Gallus gallus (Chicken).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OS Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
OX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha".
RL Mol. Cell. Biol. 8:1970-1978 (1988).
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; M20386; AAA48760.1; -
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; PARTIAL.
DR PROSITE; PS00109; PROTEIN KINASE TYR; PARTIAL.
DR PROSITE; PS00011; PROTEIN KINASE DOM; PARTIAL.
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 667 POTENTIAL.
FT DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL).
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 201 214 BY SIMILARITY.
FT DISULFID 222 230 BY SIMILARITY.
FT DISULFID 226 238 BY SIMILARITY.
FT DISULFID 239 247 BY SIMILARITY.
FT DISULFID 243 255 BY SIMILARITY.
FT DISULFID 258 267 BY SIMILARITY.
FT DISULFID 271 298 BY SIMILARITY.
FT DISULFID 302 314 BY SIMILARITY.
FT DISULFID 318 333 BY SIMILARITY.
FT DISULFID 336 340 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 581 BY SIMILARITY.
FT DISULFID 569 589 BY SIMILARITY.
FT DISULFID 592 601 BY SIMILARITY.
FT DISULFID 605 627 BY SIMILARITY.
FT DISULFID 630 638 BY SIMILARITY.
FT DISULFID 634 646 BY SIMILARITY.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON TER 703 703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;

Query Match 23.1%; Score 1575; DB 1; Length 703;
Best Local Similarity 44.2%; Pred. No. 3.8e-77;
Matches 313; Conservative 111; Mismatches 254; Indels 30; Gaps 14;
Qy 8 RWGLLLALLPPGAA-----STOVCTGDMKLRLPASPETHDMLRLHYQGCQVVGNGLE 61
Db 13 RGAIVLVLLGLGVALCSAVEKKVCGQTNKLTQLGHVEDHFTSLQRMNNCEVVLNLE 72
Qy 62 LTYLPTNASLFLQDIQEVQGVYLIAHNQVPLQRLRIVRGTLQFEDNYALAVLNGD 121
Db 73 ITVYEHNRDLTFLKTIQEVAGYVLIALNMDVDIPLNQLIIRGNVLYDNSFALVLSNH 132
Qy 122 PLNNTTPTVGASPGCLRELQRLSRSTEILKGVLIQRNPOLCYQDTILWKDIFHKNQAL 181

Db 133 -MNKTQ-----GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNDIITSRK-PL 182
Qy 182 TLID-THRSRACHPCSPKCRGWSSESSEDCQSLTRTVAGGCA-RCKGPLPTOCCHQ 239
Db 183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVICAQCCGRGKVPSCCHNQ 242
Qy 240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
Db 243 CAAGCTGPRESDCLACRFRDDATCKTCPPVLVYNPTTYQMDVNPBKYSGATCVREC 302
Qy 300 PNYLSTDVSGCTLVCPLHNOEVTAEDGTQRCCKSPCARVCYGLGMOYIKANSKFIGI 359
Db 303 PHNVVVDHSGSVRSCNTDTYEV-EENGVRCKCKDGLCKVCNGICIGELKGILS-INA 360
Qy 360 TELE-FAGCKKIFGSLAPLPESFDGDPASNTAPLOPELOVPELLEBITGVLYISAWPDS 418
Db 361 TNIDSFKNCTKINGDVSILPVAFGLDAPTKTLPDPKKLDVFTVKEISGFLLIQAWPDN 420
Qy 419 LPDLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGSGLALIHNTHLFCFVH 478
Db 421 ATDLIYAFENLEIIRGRTKQHQYSLAVNVNLIQSLGLSLKEISDGDIAIMKNVLCYAD 480
Qy 479 TVPMDQLFRQYIKANSKFIGITELECVGEGEGLACHQLCARGHCWGPGTQCVCNSQFLRQ 538
Db 481 TMNWSLFAFQSQKTKIIONENKNDCTADRHVCDPLCSDVCGWCGPGPHCFSCRFESRQK 540
Qy 539 ECVEECRVLOQLPREYVNAHCLCHPECQPNQ---SVTCGPEADOCVACAHYKDPFF 595
Db 541 ECVKQCNILQGEPEFERDSKCLPCHSECLVQNSTAYNTTCGPGPDHCKMCAHFIDGPH 600
Qy 596 CVARCPGKPDLSYMPIWPKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTSIV 655
Db 601 CVKACPVGLGENDTL-VWKYADANAVCOLCHPNCTRCCKGPGLEGCP---NGSKTPSIA 656
Qy 656 SAVV-GILLVVVLGVWFGLIKRRQKIRKYTMRLLOETELVEPLTP 702
Db 657 AGVVGGLLCLVVGGLGIGLYLRRR-HIVRKRTLRLQLQERELVEPLTP 703

Search completed: July 22, 2003, 08:46:30
Job time : 20.2304 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-369-383-12

Perfect score: 6815

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.GeneSeq_101002.*
1: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	6659	97.7	1255	21	Human heregulin 2
2	6659	97.7	1255	22	Human tyrosine kin
3	6659	97.7	1255	22	HER2 transgene pla
4	6659	97.7	1255	23	Human HER2 (ErbB2)
5	6653	97.6	1255	17	HER-2/neu protein.
6	6653	97.6	1255	20	Human HER-2/neu pr
7	6653	97.6	1255	21	Human HER-2/neu pr
8	6653	97.6	1255	21	Amino acid sequenc
9	6653	97.6	1255	22	Human HER-2/neu pr
10	6653	97.6	1255	22	HER2/neu amino aci

11	6653	97.6	1255	23	AAE24067	Human Her-2 protei
12	6653	97.6	1255	23	AAE20479	Human Her-2/neu pr
13	6653	97.6	1255	23	AAE20479	Human Her-2/neu pr
14	6653	97.6	1255	23	AAU77114	Human Her-2/neu po
15	6614	97.1	1433	14	AAE20479	Sequence of c-erbB
16	6489	95.2	1223	23	AAU98923	Human breast cance
17	6336	93.0	1200	21	AAE20479	Human HER-2/neu pr
18	5869.5	86.1	1256	21	AAE20479	Rat HER-2/neu pr
19	5869.5	86.1	1256	21	AAE20479	Rat HER-2/neu pr
20	5846.5	85.8	1256	21	AAE20479	Mouse Her-2/neu pr
21	5846.5	85.8	1256	22	AAE20479	Mouse Her-2/neu pr
22	5846.5	85.8	1256	23	AAE20479	Amino acid sequenc
23	4739	69.5	919	21	AAE20479	Mouse Her-2/neu on
24	4739	69.5	919	23	AAE20479	Human HER-2/neu fu
25	3996.5	58.6	920	23	AAE20479	Mouse Her-2/neu ex
26	3996.5	58.6	926	23	AAE20479	Mouse Her-2/neu ex
27	3623	53.2	712	21	AAE20479	Human HER-2/neu fu
28	3623	53.2	712	23	AAE20479	Human HER-2/neu fu
29	3477	51.0	782	18	AAE20479	Her-2/neu extracel
30	3475	51.0	653	21	AAE20479	Extracellular HER-
31	3475	51.0	653	23	AAE20479	Human Her-2/neu on
32	3437	50.4	645	22	AAE20479	Human ErbB2 oncopr
33	3437	50.4	645	22	AAE20479	Human ErbB2 extrac
34	3372	49.5	951	21	AAE20479	DC88GFV-erbB2EC fu
35	3269	48.0	624	11	AAE20479	Extracellular port
36	3097	45.4	1210	21	AAE20479	Amino acid sequenc
37	3097	45.4	1210	21	AAE20479	Human EGF receptor
38	3097	45.4	1210	23	AAE20479	Human Her-1 protei
39	3097	45.4	1210	23	AAE20479	Human epidermal gr
40	3095	45.4	1210	22	AAE20479	Amino acid sequenc
41	3084	45.3	583	23	AAE20479	Human protein for
42	3084	45.3	587	23	AAE20479	Human protein for
43	3083	45.2	589	23	AAE20479	Human protein for
44	3083	45.2	600	23	AAE20479	Human protein for
45	3056	44.8	1210	23	ABP51768	Human epidermal gr

ALIGNMENTS

RESULT 1
AAE20479
ID AAY92620 standard; Protein; 1255 AA.
XX AAY92620;
AC AAY92620;
XX 10-AUG-2000 (first entry)
XX Human heregulin 2 (Her2).
DE Her2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX Homo sapiens.
XX Key
XX Domain
FT Location/Qualifiers
FT 1..173
FT /label= N-terminal
FT /note= "mature polypeptide"
FT 5..25
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 59..73
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 103..117
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 149..163
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 174..323

```

FT      /label= Cysteine_rich_domain
FT      210..224
FT      /label= insertion region
FT      /note= "suitable for foreign epitope insertion"
FT      250..264
FT      /label= insertion region
FT      /note= "suitable for foreign epitope insertion"
FT      324..483
FT      /label= Ligand_binding_domain
FT      325..339
FT      /label= insertion region
FT      /note= "suitable for foreign epitope insertion"
FT      369..383
FT      /label= insertion region
FT      /note= "suitable for foreign epitope insertion"
FT      465..479
FT      /label= insertion region
FT      /note= "suitable for foreign epitope insertion"
FT      484..623
FT      /label= Cysteine_rich_domain
FT      579..593
FT      /label= insertion region
FT      /note= "suitable for foreign epitope insertion"
FT      624..654
FT      /label= Transmembrane_domain
FT      632..652
FT      /label= insertion region
FT      /note= "suitable for foreign epitope insertion"
FT      653..667
FT      /label= insertion region
FT      /note= "suitable for foreign epitope insertion"
FT      655..1010
FT      /label= Tyrosine_kinase_domain
FT      661..675
FT      /label= insertion region
FT      /note= "suitable for foreign epitope insertion"
FT      695..709
FT      /label= insertion region
FT      /note= "suitable for foreign epitope insertion"
FT      710..730
FT      /label= insertion region
FT      /note= "suitable for foreign epitope insertion"
FT      1011..1235
FT      /label= C-terminal_domain
FT
FT      WO200020027-A2.
FT
FT      13-APR-2000.
FT
FT      05-OCT-1999; 99WO-DK00525.
FT
FT      05-OCT-1998; 98DK-0001261.
FT      20-OCT-1998; 98US-0105011.
FT
FT      (MSEBI-) M & E BIOTECH AS.
FT
FT      Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
FT      Gautam A, Birk P, Karlsson G;
FT
FT      WPI; 2000-349917/30.
FT      N-:SDB; AAA09455.
FT
FT      Inducing immune responses to weakly immunogenic, tumor associated
FT      peptide antigens for the treatment of breast and prostate cancer
FT
FT      Claim 62; Page 193-198; 220pp; English.
FT
FT      This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
FT      Her2 can be used in the claimed method as an autovaccine to induce a CTL
FT      response. Subdominant CTL epitopes, antibody binding regions and
FT      cysteine residues involved in disulfide bonds are preserved in the
FT      immunogenized forms. Regions suitable for the insertion of foreign T
FT      helper epitopes were identified (see features table). The method

```

```

CC      is used for inducing immune responses against weakly immunogenic
CC      cell-associated peptide antigens (PA) such as those associated with
CC      cancers (self-proteins), e.g. human prostate specific membrane antigen
CC      (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC      The method comprises effecting simultaneous presentation by antigen
CC      producing cells (APCs) of the animals immune system of: (1) at least 1
CC      CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC      B-cell group derived from the cell-associated PA; and (2) at least 1
CC      first T helper cell group which is foreign to the animal. Analogues of
CC      human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC      part of all known and predicted CTL and B-cell epitopes of the respective
CC      PA and including at least one foreign T helper epitope are also claimed.
CC      The method is used to treat prostate, prostate/breast or breast cancer
CC      when the PA is human PSM, FGF8b and Her2, respectively.
XX

```

Seq Sequence 1255 AA;

Query Match 97.7%; Score 6659; DB 21; Length 1255;
 Best Local Similarity 97.8%; Pred. No. 0;
 Matches 1227; Conservative 7; Mismatches 21; Indels 0; Gaps 0;

```

QY      1 MELAALCRWGLLLALLPFGAASCTGCTDMKRLPASPETHLDMRLHYQGVVQGNL 60
DB      1 MELAALCRWGLLLALLPFGAASCTGCTDMKRLPASPETHLDMRLHYQGVVQGNL 60

QY      61 ELTYLPTNASLSFLQDIOEVQYVLIAHNQVRQVPLQRLIRVGTQLPEDNYALAVLDNG 120
DB      61 ELTYLPTNASLSFLQDIOEVQYVLIAHNQVRQVPLQRLIRVGTQLPEDNYALAVLDNG 120

QY      121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
DB      121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHKNNQLA 180

QY      181 LTLIDTNRSRACHPCSPMKGRCWSESSEDCQSLTRTVACGGCARCGPLPTDCHEQC 240
DB      181 LTLIDTNRSRACHPCSPMKGRCWSESSEDCQSLTRTVACGGCARCGPLPTDCHEQC 240

QY      241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTFGASCVTACP 300
DB      241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTFGASCVTACP 300

QY      301 YNYLSTDVGSCTLVCPLNQVETADGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360
DB      301 YNYLSTDVGSCTLVCPLNQVETADGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360

QY      361 ELEFAGCKKIIFGSLAFIPESFDGDPASNTAPQYIKANSKFGITELTGLYIISAWPDSLP 420
DB      361 IOEFAGCKKIIFGSLAFIPESFDGDPASNTAPQYIKANSKFGITELTGLYIISAWPDSLP 420

QY      421 DLSVFQNLQVIRGRILHNGAYSILTLOGLGISWLGLRSURELGSGLALIHNTHLFCFVHTV 480
DB      421 DLSVFQNLQVIRGRILHNGAYSILTLOGLGISWLGLRSURELGSGLALIHNTHLFCFVHTV 480

QY      481 PWDQLFRNPFOALLHTANRPEDECVGEGLAGHQLCARGHCWGPPTQCVCNSQFIRGQBC 540
DB      481 PWDQLFRNPFOALLHTANRPEDECVGEGLAGHQLCARGHCWGPPTQCVCNSQFIRGQBC 540

QY      541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600
DB      541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600

QY      601 PSGVKPDLISYMPIWKFPDEGACQPCPINCTHSCVDLDDKGPAPQASPLTISIVSAVVG 660
DB      601 PSGVKPDLISYMPIWKFPDEGACQPCPINCTHSCVDLDDKGPAPQASPLTISIVSAVVG 660

QY      661 ILLVVVLGWVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
DB      661 ILLVVVLGWVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

QY      721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780
DB      721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780

```


QY 721 RKVKVLGSGAGFTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKIILDEAYVMAGVGP 780
DB 721 RKVKVLGSGAGFTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKIILDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVTQIMPGCLLDHVRNRRGLSGQDLLNWCQIAKGSYLEDVR 840
DB 781 YVSRLLGICLTSTVQLVTQIMPGCLLDHVRNRRGLSGQDLLNWCQIAKGSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRFT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYIMVKCW 960
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYIMVKCW 960
QY 961 IDSECRPRELVSEFSRARDPQREVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVA 1020
DB 961 IDSECRPRELVSEFSRARDPQREVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVA 1020
QY 1021 EYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
DB 1021 EYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVPDGLGMAAKGLQSLPTHDPSPLORYSEDPTVPLPSTDGYPVAPLTCSPQPEYV 1140
DB 1081 AGSDVPDGLGMAAKGLQSLPTHDPSPLORYSEDPTVPLPSTDGYPVAPLTCSPQPEYV 1140
QY 1141 NQPDVRRPQPSREGPLPAARPAAGATLERAKTLPCKNGVVKDVFAGGAVENPEYLTQ 1200
DB 1141 NQPDVRRPQPSREGPLPAARPAAGATLERAKTLPCKNGVVKDVFAGGAVENPEYLTQ 1200
QY 1201 GGAAPQHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGDLVPV 1255
DB 1201 GGAAPQHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGDLVPV 1255

RESULT 3
ID AAB60167
XX AAB60167 standard; Protein; 1255 AA.
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
OS Homo sapiens.
OS Synthetic.
PN
PN WO200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
XX (GETH) GENENTECH INC.
PA
PI Erickson S, Schwall R;
XX
XX WPI; 2001-061962/07.
DR
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX

PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;

Query Match 97.7%; Score 6659; DB 22; Length 1255;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1227; Conservative 7; Mismatches 21; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLYQGCVVQGNL 60
DB 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNVQVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNVQVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLRELQRLSLEILKGGVLIQORNPOLCYODTILMKDIFHKNOLA 180
DB 121 DPLNNTTPVTGASPGGLRELQRLSLEILKGGVLIQORNPOLCYODTILMKDIFHKNOLA 180
QY 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTYCAGGCARCKGPLPTDCCHEQC 240
DB 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTYCAGGCARCKGPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHNOEVTABDGTQRCCKSPCARVCYGLGMQVYIKANSKFIGIT 360
DB 301 YNYLSTDVGSCTLVCPHNOEVTABDGTQRCCKSPCARVCYGLGMQVYIKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAFPLPESFDGDPASNTAPQYIKANSKFIGITELTYLYISAWPDSL 420
DB 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPQYIKANSKFIGITELTYLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSLRSLGSLALIHNTILCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSLRSLGSLALIHNTILCFVHTV 480
QY 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVCNCSQFLRQEC 540
DB 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVCNCSQFLRQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFPCVARC 600
DB 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFPCVARC 600
QY 601 PSGVKPDLSPYMPIWPFDEBEGACQPCINCTHSCVDLDDKGCAPABORASPLTSIVSAVVG 660
DB 601 PSGVKPDLSPYMPIWPFDEBEGACQPCINCTHSCVDLDDKGCAPABORASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAQMRILKETEL 720
DB 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAGFTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKIILDEAYVMAGVGP 780
DB 721 RKVKVLGSGAGFTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKIILDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVTQIMPGCLLDHVRNRRGLSGQDLLNWCQIAKGSYLEDVR 840
DB 781 YVSRLLGICLTSTVQLVTQIMPGCLLDHVRNRRGLSGQDLLNWCQIAKGSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRFT 900

Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALLESILRRPT 900
 QY 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLLEKGERLPPOPPICITIDVYIMVKWM 960
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLLEKGERLPPOPPICITIDVYIMVKWM 960
 QY 961 IDSECRPRFRELVSFEFSMARDPQRFVVIQNEIDLGSPASPLDSTFYRSILLEDGMDLVDA 1020
 Db 961 IDSECRPRFRELVSFEFSMARDPQRFVVIQNEIDLGSPASPLDSTFYRSILLEDGMDLVDA 1020
 QY 1021 EYILVPOQGFCCPDAPAGAGMWHHRSSSTRSGGDLTLGLSEEEAPRSLAPSEG 1080
 Db 1021 EYILVPOQGFCCPDAPAGAGMWHHRSSSTRSGGDLTLGLSEEEAPRSLAPSEG 1080
 QY 1081 AGSDVFDGDLGMAKGLQSLPTDHPSPLOQYSESDPTVPLPSETDGYVAPLTCSPQPYV 1140
 Db 1081 AGSDVFDGDLGMAKGLQSLPTDHPSPLOQYSESDPTVPLPSETDGYVAPLTCSPQPYV 1140
 QY 1141 NOPDVRPOPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
 Db 1141 NOPDVRPOPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
 QY 1201 GGAAPQHPHPPAFSPADNLVYWDQPPERGAPSTFKGTPTAENPEYLGLDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPADNLVYWDQPPERGAPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 4

AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 AC AAU74545;
 XX

23-APR-2002 (first entry)

Human HER2 (ErbB2) polypeptide.

Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 glial disorder; astrocytoma; hypopharyngeal carcinoma; epithelial
 disorder; glandular disorder; macroglossia; inflammatory disorder;
 stromal disorder; blastocoele; inflammatory disorder;
 angiogenic disorder; immunological disorder.

XX Homo sapiens.

XX US2002001587-A1.

XX 03-JAN-2002.

XX 16-MAR-2001; 2001US-0811123.

XX 16-MAR-2000; 2000US-189844P.

XX 05-OCT-2000; 2000US-236327P.

XX (ERIC/) ERICKSON S.

XX (SCHW/) SCHWALL R.

XX (SLIW/) SLIWOWSKI M.

XX Erickson S, Schwall R, Sliwowski M;

XX WPI; 2002-163686/21.

XX N-PSDB; ABK14058.

XX Treating tumour characterised by overexpression of epidermal growth

PT factor receptor, ErbB or cancer in mammal, comprises administering

PT anti-ErbB antibody-maytansinoid conjugate to the mammal

XX Example 3; Fig 7; 93pp; English.

XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytoma, hypopharyngeal, glandular, macrophagal,
 CC epithelial, stromal, blastocoele, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 XX polypeptide of the invention.

Sequence 1255 AA;

Query Match
 Best Local Similarity 97.7%; Score 6659; DB 23; Length 1255;
 Matches 1227; Conservative 7; Mismatches 21; Indels 0; Gaps 0;
 QY 1 MELAALCRWGLLLALLPFGAASSTQVCTGCTDMKRLRPASPETHLDMRLHYQCCVQVQNL 60
 Db 1 MELAALCRWGLLLALLPFGAASSTQVCTGCTDMKRLRPASPETHLDMRLHYQCCVQVQNL 60
 QY 61 ELTYLPTNASLSFLQDIOEVQGVYLIHNVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIOEVQGVYLIHNVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
 QY 121 DPLNNTTPVTGASPGGLRELQRLSITLTKGVLQIQRNPOLCYQDTILWKDIFHKNNQLA 180
 Db 121 DPLNNTTPVTGASPGGLRELQRLSITLTKGVLQIQRNPOLCYQDTILWKDIFHKNNQLA 180
 QY 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSILTRTVACGACARCKGPLETDCHEOC 240
 Db 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSILTRTVACGACARCKGPLETDCHEOC 240
 QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPHLNQVETADGTORCEKSKPCARVCYGLGMOYIKANSKFIGIT 360
 Db 301 YNYLSTDVGSCTLVCPHLNQVETADGTORCEKSKPCARVCYGLGMSHREVRAVTSAN 360
 QY 361 ELEPAGCKIFGSLAFIPESFDGDPASNTAPQYIKANSKFITELTGLYISAMPDLSLP 420
 Db 361 IQEPAGCKIFGSLAFIPESFDGDPASNTAPLQEPQLQVFTLEITGLYISAMPDLSLP 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLRELGLALIHNNTHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLRELGLALIHNNTHLCFVHTV 480
 QY 481 PWDOLFENPHOALLHTANRPEDECVGEGGLACHQICARGHCWGPPTQCVNCSQFLRGQEC 540
 Db 481 PWDOLFENPHOALLHTANRPEDECVGEGGLACHQICARGHCWGPPTQCVNCSQFLRGQEC 540
 QY 541 VECEVTLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPPEFCVARC 600
 Db 541 VECEVTLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPPEFCVARC 600
 QY 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSIVS VVG 660
 Db 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSIVS VVG 660
 QY 661 ILVVLVGVVFGILIKRQOKIRKYTNRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 Db 661 ILVVLVGVVFGILIKRQOKIRKYTNRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 QY 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKETLDEAYVNAVGVSP 780
 Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKETLDEAYVNAVGVSP 780
 QY 781 YVSRLLIGICLTSTVQLVTQMPYGCCLLDHVRENRRIGLSQDQLLWNCMQIAGMSYLEVDV 840

Db 781 VVRLGLICTSTVQLVTLMPYGLLDHVRNRRGLSGDILLNMQIAKMSYLEDDR 840
QY 841 LVHRLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRPT 900
Db 841 LVHRLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRPT 900
QY 901 HQSDVWSGVTVWELMTFGAKYDGIIPAREIPDLLEKGERLPQPICTIDVTMIMVKCM 960
Db 901 HQSDVWSGVTVWELMTFGAKYDGIIPAREIPDLLEKGERLPQPICTIDVTMIMVKCM 960
QY 961 IDSECRPRFRELVSFSEMRARDPQRFVVIQNEDELGPASPLDSTFYRSLLDDMDGLDVA 1020
Db 961 IDSECRPRFRELVSFSEMRARDPQRFVVIQNEDELGPASPLDSTFYRSLLDDMDGLDVA 1020
QY 1021 EYLVPQOQFFCPDPAPGAGGVHHRSSSTRSGGDLTLGLESEEEAPRSLAPSEG 1080
Db 1021 EYLVPQOQFFCPDPAPGAGGVHHRSSSTRSGGDLTLGLESEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAKQLSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEV 1140
Db 1081 AGSDVFDGDLGMAAKQLSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEV 1140
QY 1141 NOPDVRPPSPREGPLPAARPAATLERAKTLPSPKNGVVKDVPFAGGAVENPEYLTPO 1200
Db 1141 NOPDVRPPSPREGPLPAARPAATLERAKTLPSPKNGVVKDVPFAGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAFSPAFONLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1201 GGAAPQHPPPAFSPAFONLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 5

AAW01111
ID AAW01111 standard; Protein; 1255 AA.
AC AAW01111;
DT 01-JAN-1997 (first entry)
DE HER-2/neu protein.
XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovarian cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX Homo sapiens.
FH Key Location/Qualifiers
FT Domain 676..1255
FT /label= Intracellular domain
FT /note= "claimed domain, useful for immunisation"
XX
PN W09630514-A1.
XX
XX 03-OCT-1996.
XX
XX 28-MAR-1996; 96WO-US01689.
XX
XX 31-MAR-1995; 95US-0414417.
XX (UNIW) UNIV WASHINGTON.
PA Cheever MA, Disis MB;
XX
XX WPI: 1996-455361/45.
DR N-PSDB; AAT40739.
XX
XX DNA encoding HER-2-neu poly;peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
PS Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 97.6%; Score 6653; DB 17; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1255; Conservative 8; Mismatches 22; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPFGAASSTQVCTGTDMLKRLPASPTHLDMLRLHYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPFGAASSTQVCTGTDMLKRLPASPTHLDMLRLHYQGCVVQGNL 60
QY 61 ELYLPTNASLSFLQDIQEVQGYVLIANNQVRQVPLQRLRIVRGTLQDPEDNYALAVLDNG 120
Db 61 ELYLPTNASLSFLQDIQEVQGYVLIANNQVRQVPLQRLRIVRGTLQDPEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLREQLRLSLEILKGGVLIQRNPOLCVQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPVTGASPGGLREQLRLSLEILKGGVLIQRNPOLCVQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNSRACHPCSPMKSGSRGWESSEDCQSLTRTVACGCGARCKGPLPTDCCHEQC 240
Db 181 LTLIDTNSRACHPCSPMKSGSRGWESSEDCQSLTRTVACGCGARCKGPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTAC 300
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTAC 300
QY 301 YNYLSTDVGSCTLVCLPHNQEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCLPHNQEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAFPESEFSDGDPASNTAPQYIKANSKFIGITELTGLYISAWPDSL 420
Db 361 IQEPAGCKKIFGSLAFPESEFSDGDPASNTAPLQPEQLQVFETLEEITGLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRSLRELGSGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRSLRELGSGLALIHNTLHLCFVHTV 480
QY 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHOLCARHCWGPPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHOLCARHCWGPPTQCVNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKPPFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKPPFCVARC 600
QY 601 PSQVXPDLSPYMPIWKFPDEEGACQPCINCHTSCVDLDDKGPAPQASPLTSIVSAVVG 660
Db 601 PSQVXPDLSPYMPIWKFPDEEGACQPCINCHTSCVDLDDKGPAPQASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQRIILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQRIILKETEL 720
QY 721 RKVKVLSGAFCTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGS 780
Db 721 RKVKVLSGAFCTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGS 780
QY 781 YVSRLLGLICTSTVQLVTLMPYGLLDHVRNRRGLSGDILLNMQIAKMSYLEDDR 840
Db 781 YVSRLLGLICTSTVQLVTLMPYGLLDHVRNRRGLSGDILLNMQIAKMSYLEDDR 840

QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVYMWKCM 960
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVYMWKCM 960
QY 961 IDSECRPRELVSEFSRMRDQRFVVIQNEEDLGPASPLDSTFYRSLLLEDDMGDLVDA 1020
DB 961 IDSECRPRELVSEFSRMRDQRFVVIQNEEDLGPASPLDSTFYRSLLLEDDMGDLVDA 1020
QY 1021 EBYLVPOQGFPCDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
DB 1021 EBYLVPOQGFPCDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVLPSETDGYVAPLTCSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGGAVENPEYLTQ 1200
DB 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGGAVENPEYLTQ 1200
QY 1201 GGAAPQHPHPPAFSPAFDNLYYWDDOPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
DB 1201 GGAAPQHPHPPAFSPAFDNLYYWDDOPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 7

ID AAB21198
AAAB21198 standard; protein; 1255 AA.

AC AAB21198;

DT 12-JAN-2001 (first entry)

DE Human HER-2/neu protein.

KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;

KW breast cancer; prostate cancer; ovarian cancer; lung cancer;

KW colon cancer.

OS Homo sapiens.

PN WO200044899-A1.

PD 03-AUG-2000.

PF 28-JAN-2000; 2000WO-US02164.

PR 29-JAN-1999; 99US-0117976.

PA (CORI-) CORIXA CORP.

PA (SMIK) SMITHKLINE BEECHAM.

PI Cheever MA, Gheysen D;

DR WPI: 2000-505976/45.

DR N-PSDB; AAA89736.

XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -

XX Claim 52; Fig 7; 128pp; English.

CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.

XX

SQ Sequence 1255 AA;

Query Match 97.6%; Score 6653; DB 21; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPGCAASTQVCTGDMKRLRPASPETHLDMLRHLYQGCQVQGNL 60

DB 1 MELAALCRWGLLLALLPGCAASTQVCTGDMKRLRPASPETHLDMLRHLYQGCQVQGNL 60

QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNQVRQVPLQRLRIVRGTQOLFEDNYALAVLDNG 120

DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNQVRQVPLQRLRIVRGTQOLFEDNYALAVLDNG 120

QY 121 DPLNNTTPTVGASPGGLRELQRLSRLTEILKGGVLIQORNQVLTILWKDIFHKQNQLA 180

DB 121 DPLNNTTPTVGASPGGLRELQRLSRLTEILKGGVLIQORNQVLTILWKDIFHKQNQLA 180

QY 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGGCARCKGPLPTDCCHEOC 240

DB 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGGCARCKGPLPTDCCHEOC 240

QY 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300

DB 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300

QY 301 YNYLSTDVSGSCTLVCPLNHQNVEVTABDGTORCKSKPCARVCYGLGMQYIKANSKFIGIT 360

DB 301 YNYLSTDVSGSCTLVCPLNHQNVEVTABDGTORCKSKPCARVCYGLGMHREVRVTSAN 360

QY 361 ELEFAGCKKIFGSLAFPLESFDGDPASNTAPOYIKANSKFIGITELTGVLYSAMDLSLP 420

DB 361 IQEFAGCKKIFGSLAFPLESFDGDPASNTAPQLQVLFETLEEITGLYISAMDLSLP 420

QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSGLALIHNTHLCPVHTV 480

DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSGLALIHNTHLCPVHTV 480

QY 481 PWDQLFRNPHQALLHTANRPEDECYEGGLACHQLCARGHCWPGPTQCVNCSQFLRGQEC 540

DB 481 PWDQLFRNPHQALLHTANRPEDECYEGGLACHQLCARGHCWPGPTQCVNCSQFLRGQEC 540

QY 541 VEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFPCVARC 600

DB 541 VEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFPCVARC 600

QY 601 PSGVKPDLISYMPIWKFPDEEGACQPCINCTHSCVDLDDKCPAEORASPLTSIVSAVVG 660

DB 601 PSGVKPDLISYMPIWKFPDEEGACQPCINCTHSCVDLDDKCPAEORASPLTSIVSAVVG 660

QY 661 ILLVVVLGVVFGILIKRRQOKIRKYMRLRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

DB 661 ILLVVVLGVVFGILIKRRQOKIRKYMRLRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

QY 721 RKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780

DB 721 RKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780

QY 781 YVSRLLGICLTSTVOLVTQLMPYGLDHRVNRGRGLSGODLLNMCQIAKMSYLEDVR 840

DB 781 YVSRLLGICLTSTVOLVTQLMPYGLDHRVNRGRGLSGODLLNMCQIAKMSYLEDVR 840

QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRRT 900

DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRRT 900

QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVYMWKCM 960

DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVYMWKCM 960

QY 961 IDSECRPRELVSEFSRMRDQRFVVIQNEEDLGPASPLDSTFYRSLLLEDDMGDLVDA 1020

Db 961 IDSECRPRFRELSEFSRMDPQRFVVIQNEDELGPASPLDSTFYRSLLDDDDGLVDA 1020
Qy 1021 EYLVPQOGFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPQOGFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMGAAGLQSLPTHTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMGAAGLQSLPTHTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVPFAGGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERPKTLPSPKNGVVKDVPFAGGAVENPEYLTPO 1200
Qy 1201 GGAAPOHPHPPAFSPADNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1201 GGAAPOHPHPPAFSPADNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 8

AA584780
ID AAY84780 standard; Protein; 1255 AA.

AC AAY84780;

DT 08-AUG-2000 (first entry)

DE Amino acid sequence of the SPLICE erbb-2 receptor protein.

KW SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.

OS Homo sapiens.

PN W0200020579-A1.

PD 13-APR-2000.

PF 01-OCT-1999; 99MO-CA00912.

PR 02-OCT-1998; 98US-0165192.

PA (UYMC-) UNIV MCMASTER.

PI Muller WJ, Siegel PM;

DR WPI; 2000-303768/26.

DR N-PSDB; AAA14812.

PT Nucleic acid encoding an erbb 2 receptor protein designated SPLICE
erbb-2, inhibitors of the protein are useful for treatment of cancer -

PS Claim 3; Fig 2; 60pp; English.

CC The present sequence represents a SPLICE erbb-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbb-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbb-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbb-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbb-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.

XX Sequence 1255 AA;

Query Match 97.6%; Score 6653; DB 21; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPPGAASSTOCTCTDMKRLRASPETHLDMRLHYQCCOVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTOCTCTDMKRLRASPETHLDMRLHYQCCOVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNOVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNOVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWSESSEDCQSLTRTVACGACARCKPLPTDCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSRGWSESSEDCQSLTRTVACGACARCKPLPTDCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFILPESFDGDPASNTAPQYIKANSKFITGTLTYLISAWPDSLIP 420
Db 361 IQEFAGCKKIFGSLAFILPESFDGDPASNTAPLOPELQVFPETLEEITGYLISAWPDSLIP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSITLQGLIGISWLGRLSRELGSLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSITLQGLIGISWLGRLSRELGSLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFRNPHOALLHTANRPEDECVGEGLAHQCHGCHGPGPTQCVNCSQFURGOEC 540
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGLAHQCHGCHGPGPTQCVNCSQFURGOEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTFCFPEADQCVACAHYKDPDFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTFCFPEADQCVACAHYKDPDFCVARC 600
Qy 601 PSGVKPDLSTYMPIWKFPDEGACOPCINCTHSCVDLDDKCPABORASPLTSIISAVVG 660
Db 601 PSGVKPDLSTYMPIWKFPDEGACOPCINCTHSCVDLDDKCPABORASPLTSIISAVVG 660
Qy 661 ILLVVVLGVVFGIILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGIILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAVGSP 780
Db 721 RKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAVGSP 780
Qy 781 YVSRLLIGICTSTVOLVTLQMPYGLLDHVRNRRGLSGODLLNKCQIAKMSVLEDRV 840
Db 781 YVSRLLIGICTSTVOLVTLQMPYGLLDHVRNRRGLSGODLLNKCQIAKMSVLEDRV 840
Qy 841 LVHRDLAARNVVKSPNHVKITDFGLARLLIDIDETEHADGKVPKIKMALESILRRRPT 900
Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLIDIDETEHADGKVPKIKMALESILRRRPT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDYVIMVKCM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDYVIMVKCM 960
Qy 961 IDSECRPRFRELSEFSRMDPQRFVVIQNEDELGPASPLDSTFYRSLLDDDDGLVDA 1020
Db 961 IDSECRPRFRELSEFSRMDPQRFVVIQNEDELGPASPLDSTFYRSLLDDDDGLVDA 1020
Qy 1021 EYLVPQOGFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080

Db	1021	EEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEARPSLAPSEG	1080
Qy	1081	AGSDVFDGDLGMAAGLQSLPHDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMAAGLQSLPHDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTQ	1200
Db	1141	NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTQ	1200
Qy	1201	GGAAQPPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP	1255
Db	1201	GGAAQPPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP	1255
RESULT 9			
AAB85458			
ID	AAB85458 standard; Protein; 1255 AA.		
XX	AAB85458;		
AC			
DT	25-SEP-2001 (first entry)		
XX	Human HER-2/neu protein.		
DE			
XX	Antigen-presenting cell; immunogenic; immune response; HER-2/neu;		
KW	oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200153463-A2.		
XX			
PD	26-JUL-2001.		
XX			
PF	19-JAN-2001; 2001WO-US01850.		
XX			
PR	21-JAN-2000; 2000US-0177545.		
XX			
PA	(CORI-) CORIXA CORP.		
XX			
PI	Cheever MA, Hand-Zimmermann S;		
XX			
DR	WPI; 2001-476112/51.		
XX	N-PSDB; AAH23392.		
PT	New antigen-presenting cells, useful as vaccines for eliciting or		
PT	enhancing an immune response to HER-2/neu protein, particularly useful		
PT	for treating or preventing cancer, e.g. breast cancer -		
XX			
PS	Claim 2; Page 41-46; 49pp; English.		
XX			
CC	The invention provides an isolated antigen-presenting cell, which		
CC	expresses at least an immunogenic portion of a polypeptide that produces		
CC	an immune response to HER-2/neu protein. The antigen-presenting cells are		
CC	useful as vaccines for eliciting or enhancing an immune response to		
CC	HER-2/neu protein, particularly in treating or preventing malignancies in		
CC	which the HER-2/neu oncogene is associated. Specifically, these are		
CC	useful for treating or preventing cancer, e.g. breast cancer, ovarian,		
CC	colon, lung or prostate cancers. The present sequence represents		
CC	the human HER-2/neu protein (also known as p185 or c-erbB2).		
XX			
SQ	Sequence 1255 AA;		
Query Match			
Best Local Similarity 97.6%; Score 6653; DB 22; Length 1255;			
Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;			
Qy	1	MELAAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPEETHDMLRHLHYQCQVQGNL	60
Db	1	MELAAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPEETHDMLRHLHYQCQVQGNL	60
Qy	61	ELTYLPTNASLSFLQDIEVQGVYLIAHNQVRQVPLQRLRIVRGTLFEDNYALAVLDNG	120

Db	61	ELTYLPTNASLSFLQDIEVQGVLIAHNQVRQVPLQRLRIVRGTLFEDNYALAVLDNG	120
Qy	121	DFLNTTPTVGTASPGRLRLQSLTEILKGGVLIQRNPOLCYQDTILWKDIFPHKNOLA	180
Db	121	DFLNTTPTVGTASPGRLRLQSLTEILKGGVLIQRNPOLCYQDTILWKDIFPHKNOLA	180
Qy	181	LTLIDTNSRACHPCSPCKGSRCKGSEDECSLSTRVCAGGCARCKGKPLPTDCCHEOC	240
Db	181	LTLIDTNSRACHPCSPCKGSRCKGSEDECSLSTRVCAGGCARCKGKPLPTDCCHEOC	240
Qy	241	AAGCTGPKHSDCLACLFHNSGICBLHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP	300
Db	241	AAGCTGPKHSDCLACLFHNSGICBLHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP	300
Qy	301	YNYLSTDVGSCTLVCPHNOEVTABDGTORCEKSKPCARVCYGLGQWYIKANSKFIGIT	360
Db	301	YNYLSTDVGSCTLVCPHNOEVTABDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN	360
Qy	361	ELEFAGCKKI FGSFLAPLPESFGDPPASNTAPQYIKANSKFIGITELTGYLYISAWDPSLP	420
Db	361	IOEPAGCKKI FGSFLAPLPESFGDPPASNTAPLOPEQLQVFPETLEEITGYLYISAWDPSLP	420
Qy	421	DLVSFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRSLRGLALIHNNTHLCFVHTV	480
Db	421	DLVSFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRSLRGLALIHNNTHLCFVHTV	480
Qy	481	PHDQLFRPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVCNCSQFLRGQEC	540
Db	481	PHDQLFRPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVCNCSQFLRGQEC	540
Qy	541	VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCTCFGEADQCVACAHYKOPPCVCARC	600
Db	541	VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCTCFGEADQCVACAHYKOPPCVCARC	600
Qy	601	PSGVKPDLSYMPIMKFPDDEGACQPCINCTHSCVDLDDKGPAPAEORASPLTSIVSAVVG	660
Db	601	PSGVKPDLSYMPIMKFPDDEGACQPCINCTHSCVDLDDKGPAPAEORASPLTSIVSAVVG	660
Qy	661	ILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAOMRILKETEL	720
Db	661	ILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAOMRILKETEL	720
Qy	721	RKVVLGSGAFCTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYMAGVGSP	780
Db	721	RKVVLGSGAFCTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYMAGVGSP	780
Qy	781	YVSRLLGICLTSTVOLVTQLMPYGCLLDHVRENRRGLSQDLLNWCMIKAGMSYLEDVR	840
Db	781	YVSRLLGICLTSTVOLVTQLMPYGCLLDHVRENRRGLSQDLLNWCMIKAGMSYLEDVR	840
Qy	841	LVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT	900
Db	841	LVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT	900
Qy	901	HQSDVWSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGBRLPQPPCTIDVTVMYKCMW	960
Db	901	HQSDVWSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGBRLPQPPCTIDVTVMYKCMW	960
Qy	961	IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGASPPLDSTFYRSLLEDMDGLVDA	1020
Db	961	IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGASPPLDSTFYRSLLEDMDGLVDA	1020
Qy	1021	BEYLVPOQGFPCDPAPGAGMVHHRHRSSTRSGGDLTLGLEPSEEARPSLAPSEG	1080
Db	1021	BEYLVPOQGFPCDPAPGAGMVHHRHRSSTRSGGDLTLGLEPSEEARPSLAPSEG	1080
Qy	1081	AGSDVFDGDLGGAAGLQSLTHDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGGAAGLQSLTHDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTLSFGKNGVWVDVFAFGGAVENPEYLTQ	1200

Db	1141	NQPDVRQPPSPREGPLPAARPAAGATUERPKTILSPGNKGVKQVFAEGGAVENPEYLTQP	1250
Qy	1201	GGAAPQHPHPAPSPAFDNLYWDDPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255
Db	1201	GGAAPQHPHPAPSPAFDNLYWDDPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255
RESULT 10			
AAG88267	ID	AAG88267 standard; Protein; 1255 AA.	
XX	AC	AAG88267;	
XX	XX	11-SEP-2001 (first entry)	
DT	XX	HER2/neu amino acid sequence.	
DE	DE		
XX	XX	Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;	
KW	XX	immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;	
KW	XX	tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.	
OS	XX	Homo sapiens.	
XX	XX	W0200141787-A1.	
PN	XX	14-JUN-2001.	
PD	XX	11-DEC-2000; 2000WO-US33591.	
XX	XX	10-DEC-1999; 99US-0458299.	
PF	XX	(EPIM-) EPIMUNE INC.	
PR	XX	Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;	
XX	XX	Keogh E;	
PA	XX	WPI; 2001-374995/39.	
XX	XX	An isolated prepared HER2/neu epitope useful in a vaccine for inducing	
PT	XX	cellular immune responses for the prevention and treatment of cancer -	
PT	XX	Disclosure; Page 15; 1999p; English.	
PS	XX	The present invention describes isolated prepared HER2/neu epitopes (I).	
XX	XX	Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is	
CC	CC	culture in vitro and binds to a complex of an epitope (I), bound to a	
CC	CC	human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (1)	
CC	CC	and a second epitope and the peptide is less than 50 contiguous amino	
CC	CC	acids that have 100% identity with a native peptide sequence of HER2/neu;	
CC	CC	(3) a vaccine composition (III) comprising (II) and a pharmaceutical	
CC	CC	excipient; (4) an isolated nucleic acid encoding a peptide comprising	
CC	CC	(1); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic	
CC	CC	and immunostimulant activities, and can be used in vaccines. (I), (II)	
CC	CC	and (III) are useful for inducing cellular immune responses for the	
CC	CC	prevention and treatment of cancer. (I) and (II) are useful for	
CC	CC	monitoring or evaluating an immune response to a tumour-associated	
CC	CC	antigen when incubated with a T lymphocyte sample from a patient and	
CC	CC	detecting the presence of bound T lymphocyte to (I) or (II). Epitope	
CC	CC	based vaccines mean that immunosuppressive epitopes that may be present	
CC	CC	in whole antigens may be avoided. Selected epitopes may be combined to	
CC	CC	enhance immunogenicity. The possible pathological side effects caused by	
CC	CC	infectious agents or whole protein antigen is eliminated. The vaccine	
CC	CC	provides the ability to direct and focus an immune response to multiple	
CC	CC	selected antigens from the same pathogen. Epitope-based anti-tumour	
CC	CC	vaccines provides the opportunity to combine epitopes derived from	
CC	CC	multiple tumour-associated molecules addressing the problem of tumour-	
CC	CC	tumour variability and reducing the likelihood of tumour escape due to	
CC	CC	antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in	
CC	CC	the exemplification of the present invention.	
XX	XX	Sequence 1255 AA;	
SQ			

Db	1021	EEYLVPOQFFCPDPAPGAGGMVHHRRSSSTRSGGDLTLGLEPSEEBAPSLAPSEG	1080
Qy	1081	AGSDVFDGDLGCAAGLQSLTHDPSPLORYSEPTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGCAAGLQSLTHDPSPLORYSEPTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVRPQPPSPREGPLPAARPAAGATLERPKTSLPGKNGVVKDVFAFGGAVENPEYLTQ	1200
Db	1141	NQPDVRPQPPSPREGPLPAARPAAGATLERPKTSLPGKNGVVKDVFAFGGAVENPEYLTQ	1200
Qy	1201	GGAAPQHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKTPTAENPEYLGLDVPV	1255
Db	1201	GGAAPQHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKTPTAENPEYLGLDVPV	1255
RESULT 11			
AAE24067			
ID	AAE24067	standard; Protein; 1255 AA.	
XX	XX		
AC	AAE24067;		
XX	XX		
DT	23-SEP-2002	(first entry)	
XX	XX		
DE	Human Her-2 protein.		
XX	XX		
KW	Human; Her-2; epidermal growth factor receptor 2; infection; cancer;		
KW	hyperproliferative disorder; prophylaxis; inflammation; antisease;		
KW	tumour; gene therapy; phosphorothioate backbone.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
PN	WO200222636-A1.		
XX	XX		
PD	21-MAR-2002.		
XX	XX		
PF	12-SEP-2001; 2001WO-US28572.		
XX	XX		
PR	15-SEP-2000; 2000US-0663834.		
XX	XX		
PA	(ISIS-) ISIS PHARM INC.		
XX	XX		
PI	Bennett CF, Cowsert LM;		
XX	XX		
DR	WPI; 2002-471192/50.		
DR	N-PSDB; AAD38904.		
XX	XX		
PT	Novel antisease oligonucleotide which modulates the expression of Human		
PT	Epidermal Growth Factor receptor, Her2, is useful for treating tumors		
PT	Inflammation or to prevent infection in humans -		
XX	XX		
PS	Example 13; Page 95-107; 116pp; English.		
XX	XX		
CC	The invention relates to antisease compounds targetted to a nucleic		
CC	acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)		
CC	that specifically hybridises with and inhibits the expression of Her2.		
CC	Antisease compounds of the invention are used for treating diseases or		
CC	conditions associated with Her2 such as hyperproliferative disorders		
CC	e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,		
CC	neural or cardiac cancer. They are also useful prophylactically e.g.		
CC	to prevent or delay infection, inflammation and tumour formation. The		
CC	invention is also used in gene therapy. The present sequence is human		
CC	Her-2 protein.		
XX	XX		
SQ	Sequence 1255 AA;		
Query Match			
Best Local Similarity 97.6%; Score 6653; DB 23; Length 1255;			
Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;			
Qy	1	MELAAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPEHLDMLRHLYQGCQVQGNL	60
Db	1	MELAAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPEHLDMLRHLYQGCQVQGNL	60

Qy	61	ELTYLPTNASLSFLODIOEVQGYVLIAHNQVQVPLQRLRIVRGTQLFEDNYALAVLDNG	120
Db	61	ELTYLPTNASLSFLODIOEVQGYVLIAHNQVQVPLQRLRIVRGTQLFEDNYALAVLDNG	120
Qy	121	DPLNNTTPTVGASPGGLRELQRLSUTELKGGVLQIRNPOLCYQDITLWKDIFHKQNQLA	180
Db	121	DPLNNTTPTVGASPGGLRELQRLSUTELKGGVLQIRNPOLCYQDITLWKDIFHKQNQLA	180
Qy	181	LTLIDTNSRACHPCSPMKGSRCHGESEDCOSLTRTYCAGGCARCKPLPTDCCHEOC	240
Db	181	LTLIDTNSRACHPCSPMKGSRCHGESEDCOSLTRTYCAGGCARCKPLPTDCCHEOC	240
Qy	241	AAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCYTACP	300
Db	241	AAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCYTACP	300
Qy	301	YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN	360
Db	301	YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN	360
Qy	361	ELEFAGCKKIFGSLAPLPESFDGDPASNTAPQYIKANSKFIGITELTGYLYISAWPDSL	420
Db	361	ELEFAGCKKIFGSLAPLPESFDGDPASNTAPQYIKANSKFIGITELTGYLYISAWPDSL	420
Qy	421	DLVSFQNLQVIRGRILHNGAYSLLTQGLGISWGLRSLRELGSGLALIHNTLHLCFVHTV	480
Db	421	DLVSFQNLQVIRGRILHNGAYSLLTQGLGISWGLRSLRELGSGLALIHNTLHLCFVHTV	480
Qy	481	PMDQLFRNPQHALLHTANRPEDECYEGGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC	540
Db	481	PMDQLFRNPQHALLHTANRPEDECYEGGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC	540
Qy	541	VEECRVLOGLPREYUNARHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDPFPCVARC	600
Db	541	VEECRVLOGLPREYUNARHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDPFPCVARC	600
Qy	601	PSGVKPDLSYMPIMKFPDEEGACQPCINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG	660
Db	601	PSGVKPDLSYMPIMKFPDEEGACQPCINCTHSCVDLDDKGCPAEQRASPLTSIISAVVG	660
Qy	661	ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL	720
Db	661	ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL	720
Qy	721	RKVYLVGSGAGCTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYMAGVGP	780
Db	721	RKVYLVGSGAGCTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYMAGVGP	780
Qy	781	YVSRLLGICLTSTVOLVTQLMFYGCLLDHVRENRRGLSGQDLLNWCQIAKGSYLEDVR	840
Db	781	YVSRLLGICLTSTVOLVTQLMFYGCLLDHVRENRRGLSGQDLLNWCQIAKGSYLEDVR	840
Qy	841	LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESIILRRRFT	900
Db	841	LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESIILRRRFT	900
Qy	901	HOSDVMVSGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTTIDVMYMKWM	960
Db	901	HOSDVMVSGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTTIDVMYMKWM	960
Qy	961	IDSECRPRFRELVSFSESRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDMDGLDVA	1020
Db	961	IDSECRPRFRELVSFSESRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDMDGLDVA	1020
Qy	1021	EYLVPOQFFCPDPAPGAGGMVHHRRSSSTRSGGDLTLGLEPSEEBAPSLAPSEG	1080
Db	1021	EYLVPOQFFCPDPAPGAGGMVHHRRSSSTRSGGDLTLGLEPSEEBAPSLAPSEG	1080
Qy	1081	AGSDVFDGDLGCAAGLQSLTHDPSPLORYSEPTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGCAAGLQSLTHDPSPLORYSEPTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVRPQPPSPREGPLPAARPAAGATLERPKTSLPGKNGVVKDVFAFGGAVENPEYLTQ	1200

Db 1141 NQPDVRQPPSPFRGGLPAAPAGATLERPKTLSPGKNGVVKDVFAGGAVENPEYLTPQ 1200
Qy 1201 GGAAPQHPHPPAFPAFDNLYYWDQPPPERGAPPSFTFKGTPTAENPEYLGLDVPV 1255
Db 1201 GGAAPQHPHPPAFPAFDNLYYWDQPPPERGAPPSFTFKGTPTAENPEYLGLDVPV 1255
RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX AC AAE20479;
XX AC
XX AC
DT 01-JUL-2002 (first entry)
XX DE Human Her-2/neu protein.
XX KW Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
XX KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Region 1021..1030
XX FT /note= "Naturally processed HLA-B44-restricted epitope"
XX FN WO200214503-A2.
XX PD 21-FEB-2002.
XX PF 14-AUG-2001; 2001WO-US41733.
XX PR 14-AUG-2000; 2000US-225152P.
XX PR 28-SEP-2000; 2000US-236428P.
XX PR 21-FEB-2001; 2001US-270520P.
XX PA (CORI-) CORIXA CORP.
XX PI Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
XX PI McNeill PD, Vedvick TS;
XX WPI: 2002-280758/32.
XX DR N-PSDB; AAD32743.
XX PT Novel isolated Her-2/Neu polypeptide composition useful for therapy,
XX PS prevention and diagnosis of cancer, preferably breast cancer -
XX PS Disclosure; Page 114-117; 129pp; English.
XX CC The invention relates to an isolated Her-2/Neu polypeptide composition
XX CC effective for eliciting an immune response. The invention is useful for
XX CC eliciting an immune response in a patient, where the patient is human
XX CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
XX CC The composition is useful for the therapy and diagnosis of cancer,
XX CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
XX CC and other compositions for the diagnosis, prevention and treatment of
XX CC human malignancies, for stimulating and/or expanding T cells specific for
XX CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
XX CC patient. The invention is useful for stimulating a T cell response in a
XX CC human patient, as probe or primer for nucleic acid hybridization, to
XX CC selectively form duplex molecules with complementary stretches of the
XX CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
XX CC length gene from a suitable library, and to direct expression of a
XX CC polypeptide in appropriate host cells. The composition is useful in
XX CC prophylactic or therapeutic applications and for the treatment of cancer,
XX CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
XX CC associated malignancies. The invention is useful in gene therapy. The
XX CC present sequence is human Her-2/neu protein.

Best Local Similarity 97.6%; Pred. No. 0;
Matches 1255; Conservative 8; Mismatches 22; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPPGAASSTOVCTGTDMLRLPASPEHLDMLRLHYQCCVVOGHL 60
Db 1 MELAALCRWGLLLALLPPGAASSTOVCTGTDMLRLPASPEHLDMLRLHYQCCVVOGHL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNOVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNOVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVCASPGGLREQLRSLTEILKGGVLIQRPOLCYQDTILWKDIFHKQNQLA 180
Db 121 DPLNNTTPTVCASPGGLREQLRSLTEILKGGVLIQRPOLCYQDTILWKDIFHKQNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACAGGCARCKGPLPTDCHEOC 240
Db 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACAGGCARCKGPLPTDCHEOC 240
Qy 241 AAGCTGPKHSDCLACHLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACHLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNQVTAEDGTQRCCKSPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHLNQVTAEDGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPQYIKANSKFIGITELTGLVYISAMPDSL 420
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELQVFTLEBITGLYISAMPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRLSRLGSLGIALIHNHNLFCVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRLSRLGSLGIALIHNHNLFCVHTV 480
Qy 481 PWDQLFRPHQALLHTANRPEDECVGEGSLACHQLCARGCHGPGPTQCNCVSQFIRGQEC 540
Db 481 PWDQLFRPHQALLHTANRPEDECVGEGSLACHQLCARGCHGPGPTQCNCVSQFIRGQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPEQOPQNGSVTCFGEADQCACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPEQOPQNGSVTCFGEADQCACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMIWKFPDEGACQPCPINCTHSCVDLDDKGCPEAQASPLTISIVAVVG 660
Db 601 PSGVKPDLSPYMIWKFPDEGACQPCPINCTHSCVDLDDKGCPEAQASPLTISIVAVVG 660
Qy 661 ILLVVVLGVVFGIILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGIILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTLMPYGCLLDHVRENRLGSDQLLNKCMQIAKMSYLEBVR 840
Db 781 YVSRLLGICLTSTVQLVTLMPYGCLLDHVRENRLGSDQLLNKCMQIAKMSYLEBVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDTEYHADGCKGVPKWMALLESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDTEYHADGCKGVPKWMALLESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYIMVKCM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYIMVKCM 960
Qy 961 IDSECRPRFRELVSFSEFRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1020
Db 961 IDSECRPRFRELVSFSEFRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1020
Qy 1021 EYVLVPOQGFCCPDPAPGAGMVHHRSSRSGGDLTLGLLEPSEEEAPRSPAPSE 1080
Query Match 97.6%; Score 6653; DB 23; Length 1255;

Db 1021 EEYLVPOQFFCFDPAPGAGCMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVLPSETDGYVAPLTCSPQPEV 1140
Db 1081 AGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVLPSETDGYVAPLTCSPQPEV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAFGGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAFGGAVENPEYLTPO 1200
Qy 1201 CGAAPOPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKTPTAENPEYLGLOVPV 1255
Db 1201 CGAAPOPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKTPTAENPEYLGLOVPV 1255

RESULT 13

AAM51143

ID AAM51143 standard; Protein; 1255 AA.

XX AC AAM51143;

XX DT 17-JUN-2002 (first entry)

XX DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

XX KW tyrosine kinase; receptor; c-erbB2; gene therapy.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Domain 1..653

FT Domain /note="extracellular domain"

FT Domain 676..1255

FT Domain /note="intracellular domain"

FT Domain 990..1255

FT Domain /note="phosphorylation domain"

XX WO200212341-A2.

XX PN 14-FEB-2002.

XX PD 03-AUG-2001; 2001WO-US24283.

XX PF 03-AUG-2000; 2000US-0632507.

XX PR (CORI-) CORIXA CORP.

XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX DR N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting

XX or enhancing an immune response to the protein, has Her-2/neu

XX extracellular domain fused to Her-2/neu intracellular or

XX phosphorylation domain -

XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein

XX or c-erbB2), an oncogenic self-protein and target for anti-cancer

XX vaccines. The Her-2/neu gene is amplified and p185 is overexpressed

XX in a variety of cancers, including breast, ovarian, colon, lung and

XX prostate cancer. Her-2/neu is a member of the tyrosine kinase

XX family of receptor-like glycoproteins. It comprises an extracellular

XX domain with homology to the epidermal growth factor receptor

XX (EGFR), a highly hydrophobic transmembrane domain and a C-terminal

XX intracellular domain that also shows homology to EGFR. Its

XX overexpression correlates with a poor prognosis in breast and

XX ovarian cancers. The invention provides Her-2/neu fusion

XX proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its DeltapB fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Qy Query Match 97.6%; Score 6653; DB 23; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPCAASTQVCTGDMKRLPASPETHLDMRLHLYQCCVVGNL 60

Db 1 MELAALCRWGLLLALLPPCAASTQVCTGDMKRLPASPETHLDMRLHLYQCCVVGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVRQVPLQRLRIVRGTLQFEDNALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVRQVPLQRLRIVRGTLQFEDNALAVLDNG 120

Qy 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKKNQLA 180

Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKKNQLA 180

Qy 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGACRCKGPLPTDCCHEQC 240

Db 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGACRCKGPLPTDCCHEQC 240

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPPEGRVTFGASCVTAC 300

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPPEGRVTFGASCVTAC 300

Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360

Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360

Qy 361 ELEFAGCKKIFGSLAFPESEFDGDPASNTAPQYIKANSKFIGITELTYGLYISAWPDSLP 420

Db 361 IQEFAGCKKIFGSLAFPESEFDGDPASNTAPQYIKANSKFIGITELTYGLYISAWPDSLP 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480

Qy 481 PWDQLFRPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQVCNCSQFLRGQEC 540

Db 481 PWDQLFRPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQVCNCSQFLRGQEC 540

Qy 541 VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCACAHYKPPFCVAC 600

Db 541 VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCACAHYKPPFCVAC 600

Qy 601 PSGVKPDLISYMPIKFPDEEGACQPCPINCTHSCVDLDDKGCFAQRASPLTSIVSAVVG 660

Db 601 PSGVKPDLISYMPIKFPDEEGACQPCPINCTHSCVDLDDKGCFAQRASPLTSIVSAVVG 660

Qy 661 ILLVVVLGVVFGILIKRQKQIRKYTMRLLOQETELVEPLTPSGAMPNQAQMRILKETEL 720

Db 661 ILLVVVLGVVFGILIKRQKQIRKYTMRLLOQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGP 780

Db 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGP 780

Qy 781 YVRLGICLTSTVLTQMPYGCCLLDHVRNRRGRSGQDLNWKQIAKMSYLEVDV 840
Db 781 YVRLGICLTSTVLTQMPYGCCLLDHVRNRRGRSGQDLNWKQIAKMSYLEVDV 840
Qy 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKMALESILRRFT 900
Db 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCWM 960
Qy 961 IDSECRPRFRELVSFEFSMARDDPQRFVVIQNEIDGSPASPLDSTFYRSLLLEDGMDLVDA 1020
Db 961 IDSECRPRFRELVSFEFSMARDDPQRFVVIQNEIDGSPASPLDSTFYRSLLLEDGMDLVDA 1020
Qy 1021 BEYLVPQGFPCPDPAAGAGVHHRSSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1080
Db 1021 BEYLVPQGFPCPDPAAGAGVHHRSSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1080
Qy 1081 AGSDVFDGDLGWAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPPQPEV 1140
Db 1081 AGSDVFDGDLGWAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPPQPEV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVDVFAFGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVDVFAFGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX AAU77114;
AC AAU77114;
DT 05-JUN-2002 (first entry)
XX Human Her-2/neu polypeptide.
DE Human; Her-2/neu; cytostatic; haematological malignancy; CML;
XX acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX Homo sapiens.
OS WO200213847-A2.
XX 21-FEB-2002.
XX 13-AUG-2001; 2001WO-US25408.
XX 14-AUG-2000; 2000US-0638280.
XX 28-SEP-2000; 2000US-0675904.
XX (CORI-) CORIXA CORP.
XX Gaiger A, Cheever MA, Hand-zimmermann S;
XX WPI; 2002-280741/32.
XX N-PSDB; ABK10730.
XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
XX polypeptide
PS Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 97.6%; Score 6653; DB 23; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPFGAASSTVCTGDMKRLPASPETHLDMLRLHYQCQVQGNL 60
Db 1 MELAALCRWGLLLALLPFGAASSTVCTGDMKRLPASPETHLDMLRLHYQCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQYVLIHNOVRQVPLQRLRIVRGTLFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQYVLIHNOVRQVPLQRLRIVRGTLFEDNYALAVLNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPQCYQDTILWKDIFHKQNQLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPQCYQDTILWKDIFHKQNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSCWGESSEDCQSLTRTVACAGGACRCKPLDTCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSCWGESSEDCQSLTRTVACAGGACRCKPLDTCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHNEQVTAEDGTQRCCKSPKCARVCYGLGMEHLREVRAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHNEQVTAEDGTQRCCKSPKCARVCYGLGMEHLREVRAVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFIPESPDGDPASNTAPQYIKANSKFITGLTGLYISANPDSL 420
Db 361 IQEFAGCKKIFGSLAFIPESPDGDPASNTAPQYIKANSKFITGLTGLYISANPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLISWLGSLRSLRELSGLALTHNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLISWLGSLRSLRELSGLALTHNTHLCFVHTV 480
Qy 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQCHGCHGPGPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQCHGCHGPGPTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Db 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMPIWKFFDEGACQPCPINCTHSCVDLDDKCGCAPQASPLTISIYAVVG 660
Db 601 PSGVKPDLSPYMPIWKFFDEGACQPCPINCTHSCVDLDDKCGCAPQASPLTISIYAVVG 660
Qy 661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAGTGYVKGWIIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVWAGVSP 780
Db 721 RKVKVLGSGAGTGYVKGWIIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVWAGVSP 780
Qy 781 YVSRLLGICLTSTVLTQMPYGCCLLDHVRNRRGRSGQDLNWKQIAKMSYLEVDV 840
Db 781 YVSRLLGICLTSTVLTQMPYGCCLLDHVRNRRGRSGQDLNWKQIAKMSYLEVDV 840
Qy 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKMALESILRRFT 900

Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRRFT 900
QY 901 HOSDWSYGVTVWELMTFGAKYDGIIPAREIPDLLEKGERLPPOPICTIDVYMIWKWM 960
Db 901 HOSDWSYGVTVWELMTFGAKYDGIIPAREIPDLLEKGERLPPOPICTIDVYMIWKWM 960
QY 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020
Db 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020
QY 1021 EYLVPQOGFFCPDPAPGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPQOGFFCPDPAPGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEV 1140
QY 1141 NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGGAVENPEYLTPO 1200
Db 1141 NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGDLVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGDLVPV 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
AC AAR39568;
XX AAR39568;
DT 07-FEB-1994 (first entry)
XX Sequence of c-erbB-2 tumour antigen.
DE Sequence of c-erbB-2 tumour antigen.
KW Tumour antigen; c-erbB-2; glycoprotein.
OS Homo sapiens.
XX WO9316185-A.
PN 19-AUG-1993.
PD 05-FEB-1993; 93WO-US01055.
XX 06-FEB-1992; 92US-0831967.
PR (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX Houston LL, Huston JS, Oppermann H, Ring DB;
XX WPI; 1993-272889/34.
DR N-PSDB; AAQ46083.
XX New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX Disclosure; pages 48-54; 87pp; English.
PS c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents
CC the location of a stop codon in AAQ46083.
XX Sequence 1433 AA;

Query Match 97.1%; Score 6614; DB 14; Length 1433;
Best Local Similarity 97.1%; Pred. No. 0;

Matches 1219; Conservative 10; Mismatches 26; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPGCAASTOVCTGTDMLKRLPASBETHLDMRLHLYQGVQVGNL 60
Db 1 MELAALCRWGLLLALLPGCAASTOVCTGTDMLKRLPASBETHLDMRLHLYQGVQVGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGVLIQAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGVLIQAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGCLRELQRLSLEILKGGVLQORNPOLCYQDTILWKDIFHKQNOLA 180
Db 121 DPLNNTTPVTGASPGCLRELQRLSLEILKGGVLQORNPOLCYQDTILWKDIFHKQNOLA 180
QY 181 LTLIDTNRSRACHPCSPCKGRCWGSEDECSLTRTYCAGCCKARCKPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPCKGRCWGSEDECSLTRTYCAGCCKARCKPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNHQEVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPLNHQEVTAEDGTORCEKSKPCARVCYGLGMQVIRVAVTSAN 360
QY 361 ELEFAGCKKI FGS LAF LPE SFDGDPASNTAPOYIKANSKFI G I T E L T G Y L I S A M P D S L P 420
Db 361 IOEFAGCKRI FGS LAF LPE SFDGDPASNTAPOEHLQVFTLEQITG Y L I S A M P D S L P 420
QY 421 DLSVFQNLQVIRGRILHNGAYS L T L Q G L G I S W L G R L S R L E L G S G L A L I H N H T L C F V H T V 480
Db 421 DLSVFQNLQVIRGRILHNGAYS L T L Q G L G I S W L G R L S R L E L G S G L A L I H N H T L S F V H T V 480
QY 481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQEC 540
QY 541 VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFVCARC 600
Db 541 VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFVCARC 600
QY 601 PSGVKPDLISYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660
Db 601 PSGVKPDLISYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPORASPLTSIIISAVVG 660
QY 661 ILLVVVLGVVFGILIKRROOKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRROOKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLSGAGFTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKIILDEAYMAGVGP 780
Db 721 RKVKVLSGAGFTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKIILDEAYMAGVGP 780
QY 781 YVSRLLGICLSTVOLVTQLMFYGCLLDHVRNRRGRLSQDLLNMCQAKGMSYLEDVDR 840
Db 781 YVSRLLGICLSTVOLVTQLMFYGCLLDHVRNRRGRLSQDLLNMCQAKGMSYLEDVDR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRRFT 900
QY 901 HOSDWSYGVTVWELMTFGAKYDGIIPAREIPDLLEKGERLPPOPICTIDVYMIWKWM 960
Db 901 HOSDWSYGVTVWELMTFGAKYDGIIPAREIPDLLEKGERLPPOPICTIDVYMIWKWM 960
QY 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020
Db 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020
QY 1021 EYLVPQOGFFCPDPAPGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPQOGFFCPDPAPGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080

Qy	1081	AGSDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVRPOPSPREGPLPAARPAGATLERAKTLPFGKNGVVKDVFAFGGAVENPEYLTPO	1200
Db	1141	NQPDVRPOPSPREGPLPAARPAGATLERPKTLPFGKNGVVKDVFAFGGAVENPEYLTPO	1200
Qy	1201	GGAAPQHPHPPAFSPAFDNLYYWDDPPPERGAPPSTFKGTTAENPEYLGLDVPV	1255
Db	1201	GGAAPQHPHPPAFSPAFDNLYYWDDPPPERGAPPSTFKGTTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 08:41:23
Job time : 43.9774 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:41:54 ; Search time 36.7573 Seconds

(without alignments)
4527.811 Million cell updates/sec

Title: SEQ4-632-652-12

Perfect score: 6796

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVFPV 1249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A. Geneseq 101002.*

1: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6625	97.5	1255	21	Human heregulin 2
2	6625	97.5	1255	22	Human tyrosine kin
3	6625	97.5	1255	22	HER2 transgene pla
4	6625	97.5	1255	23	Human HER2 (ErbB2)
5	6619	97.4	1255	17	AAW01111
6	6619	97.4	1255	20	AAW92406
7	6619	97.4	1255	21	Human HER-2/neu pr
8	6619	97.4	1255	21	AAW21198
9	6619	97.4	1255	22	Amino acid sequenc
10	6619	97.4	1255	22	Human HER-2/neu pr
					HER2/neu amino aci

11	6619	97.4	1255	23	AAE24067	Human Her-2 protei
12	6619	97.4	1255	23	AAE20479	Human Her-2/neu pr
13	6619	97.4	1255	23	AAE20479	Human Her-2/neu pr
14	6619	97.4	1255	23	AAU77114	Human Her-2/neu po
15	6576	96.8	1433	14	AAE20479	Sequence of c-erbB
16	6455	95.0	1223	23	AAU98923	Human breast cance
17	6302	92.7	1200	21	AAE21208	Human HER-2/neu pr
18	5861.5	86.2	1256	21	AAE21199	Rat Her-2/neu pr
19	5861.5	86.2	1256	21	AAE21144	Mouse Her-2/neu pr
20	5836.5	85.9	1256	21	AAE21206	Mouse Her-2/neu pr
21	5836.5	85.9	1256	21	AAE21206	Amino acid sequenc
22	5836.5	85.9	1256	23	AAE21151	Mouse Her-2/neu on
23	4817	70.9	919	21	AAE21203	Human HER-2/neu fu
24	4817	70.9	919	21	AAE21203	Human HER-2/neu fu
25	4067.5	59.9	920	23	AAE21148	Mouse Her-2/neu ex
26	4067.5	59.9	926	23	AAE21153	Mouse Her-2/neu ex
27	3701	54.5	712	21	AAE21204	Human HER-2/neu fu
28	3701	54.5	712	21	AAE21204	Human HER-2/neu fu
29	3554	52.3	782	18	AAW19764	Her2-GM-CSF immuno
30	3550	52.2	653	21	AAE21200	Extracellular HER-
31	3550	52.2	653	23	AAE21145	Human Her-2/neu on
32	3512	51.7	645	22	AAE20408	Human ErbB2 oncopr
33	3512	51.7	645	22	AAE20408	Human ErbB2 extrac
34	3447	50.7	951	21	AAE20493	DCScFv-erbB2EC fu
35	3344	49.2	624	11	AAE20493	Extracellular port
36	3115	45.8	1210	21	AAE20493	Amino acid sequenc
37	3115	45.8	1210	21	AAE20493	Human EGF receptor
38	3115	45.8	1210	23	AAE20493	Human Her-1 protei
39	3115	45.8	1210	23	AAE20493	Human epidermal gr
40	3113	45.8	1210	22	AAE20493	Amino acid sequenc
41	3085	45.4	583	23	AAE20483	Human protein for
42	3085	45.4	583	23	AAE20483	Human protein for
43	3083	45.4	589	23	AAE20484	Human protein for
44	3083	45.4	600	23	AAE20482	Human protein for
45	3074	45.2	1210	23	ABP51768	Human epidermal gr

ALIGNMENTS

RESULT 1

AAE20482

ID AAY92620 standard; Protein; 1255 AA.

AC AAY92620;

DT 10-AUG-2000 (first entry)

XX Human heregulin 2 (Her2).

DE Herregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;

KW self-protein; cancer; breast cancer; prostate cancer;

KW cell-associated peptide antigen; foreign epitope.

OS Homo sapiens.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Domain 1..173

FT /label= N-terminal

FT /note= "mature polypeptide"

FT Region 5..25

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT Region 59..73

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT Region 103..117

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT Region 149..163

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT Domain 174..323

FT Region /label= Cysteine_rich_domain
210..224
/label= insertion region
/note= "suitable for foreign epitope insertion"
250..264
/label= insertion region
/note= "suitable for foreign epitope insertion"
324..483
/label= Ligand_binding_domain
325..339
/label= insertion region
/note= "suitable for foreign epitope insertion"
369..383
/label= insertion region
/note= "suitable for foreign epitope insertion"
465..479
/label= insertion region
/note= "suitable for foreign epitope insertion"
484..623
/label= Cysteine_rich_domain
579..593
/label= insertion region
/note= "suitable for foreign epitope insertion"
624..654
/label= Transmembrane_domain
632..652
/label= insertion region
/note= "suitable for foreign epitope insertion"
653..667
/label= insertion region
/note= "suitable for foreign epitope insertion"
655..1010
/label= Tyrosine_kinase_domain
661..675
/label= insertion region
/note= "suitable for foreign epitope insertion"
695..709
/label= insertion region
/note= "suitable for foreign epitope insertion"
710..730
/label= insertion region
/note= "suitable for foreign epitope insertion"
1011..1235
/label= C-terminal_domain
XX
W0200020027-A2.
XX
13-APR-2000.
XX
05-OCT-1999; 99WO-DK00525.
XX
05-OCT-1998; 98DK-0001261.
PR 20-OCT-1998; 98US-0105011.
XX
(WEBI-) M & E BIOTECH AS.
XX
Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
PI Gautam A, Birk P, Karlsson G;
XX
WPI; 2000-349917/30.
DR N-PSDB; AAA09455.
XX
Inducing immune responses to weakly immunogenic, tumor associated
PT peptide antigens for the treatment of breast and prostate cancer
XX
Claim 62; Page 193-198; 220pp; English.
XX
This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
CC response. Subdominant CTL epitopes, antibody binding regions and
CC cysteine residues involved in disulfide bonds are preserved in the
CC immunogenized forms. Regions suitable for the insertion of foreign T
CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
CC cell-associated peptide antigens (PA) such as those associated with
CC cancers (self-proteins), e.g. human prostate specific membrane antigen
CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC The method comprises effecting simultaneous presentation by antigen
CC producing cells (APCs) of the animals immune system of: (1) at least 1
CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC B-cell group derived from the cell-associated PA; and (2) at least 1
CC first T helper cell group which is foreign to the animal. Analogues of
CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC part of all known and predicted CTL and B-cell epitopes of the respective
CC PA and including at least one foreign T helper epitope are also claimed.
CC The method is used to treat prostate, prostate/breast or breast cancer
CC when the PA is human PSM, FGF8b and Her2, respectively.
XX
SQ Sequence 1255 AA;
Query Match 97.5%; Score 6625; DB 21; Length 1255;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1222; Conservative 6; Mismatches 16; Indels 16; Gaps 2;
QY 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLPASPETHLDMRLHLYQGCVVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLPASPETHLDMRLHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVLLIAHQVQVPLQRLRIVRGTOLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIQEVQGVLLIAHQVQVPLQRLRIVRGTOLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQORNQPCVQDTILMKDIFHKQNQLA 180
DB 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQORNQPCVQDTILMKDIFHKQNQLA 180
QY 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCHEQC 240
DB 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMNPGRVTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMNPGRVTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPPLHNQVTAEDGTQRCCKSKPCARVCYGLQMQVYKANSKFIGIT 360
DB 301 YNYLSTDVGSCTLVCPPLHNQVTAEDGTQRCCKSKPCARVCYGLQMHLREVRVTSAN 360
QY 361 ELEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPELOVFEETLEETGVLYISAWPDSL 420
DB 361 IQEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPELOVFEETLEETGVLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGLGSLALIHNTHLCPVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGLGSLALIHNTHLCPVHTV 480
QY 481 PMDQLFRNPQHALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSOFLRGQEC 540
DB 481 PMDQLFRNPQHALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSOFLRGQEC 540
QY 541 VEECRVLOGLPREYVYNAHCLPCHPECPQNGSVTCFGEPEADQCACAHYKDPFPCVARC 600
DB 541 VEECRVLOGLPREYVYNAHCLPCHPECPQNGSVTCFGEPEADQCACAHYKDPFPCVARC 600
QY 601 PSGVXPDLISYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPORASPLTSIQVIKAN 660
DB 601 PSGVXPDLISYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPORASPLTSIV----- 655
QY 661 SKFIGI-----TELKRRQOKIRKYTMRRLQETELVEPLTPSGAMPNQAQMRIL 709
DB 656 SAVVGILLVVLVGVFGILIKRRQOKIRKYTMRRLQETELVEPLTPSGAMPNQAQMRIL 715
QY 710 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMA 769
DB 716 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMA 775

Qy 770 GVGSPPYVSRLLIGICTSTVQLVTLQMPYGCCLLDHVRENRLGSGDILLNWCWQIAKMSY 829
Db 776 GVGSPPYVSRLLIGICTSTVQLVTLQMPYGCCLLDHVRENRLGSGDILLNWCWQIAKMSY 835
Qy 830 LEDVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGGKVPKMMALESIL 889
Db 836 LEDVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGGKVPKMMALESIL 895
Qy 890 RRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIM 949
Db 896 RRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIM 955
Qy 950 VKCWMIDSECRPRELVSEFSRMARDPQRFVWQNEIDLGASPLDSTFYRSLLEDDMG 1009
Db 956 VKCWMIDSECRPRELVSEFSRMARDPQRFVWQNEIDLGASPLDSTFYRSLLEDDMG 1015
Qy 1010 DLVDAEEVLPVQGFPCPDPAAGAGMVHRRHRSSTRSGGDLTLGLEPSEEEAPRSPL 1069
Db 1016 DLVDAEEVLPVQGFPCPDPAAGAGMVHRRHRSSTRSGGDLTLGLEPSEEEAPRSPL 1075
Qy 1070 APSEGAGSDVFDGLGMAAGKLGSLPHTDPSPIQRYSEDPTVPLPSETDGYVAPLTCSP 1129
Db 1076 APSEGAGSDVFDGLGMAAGKLGSLPHTDPSPIQRYSEDPTVPLPSETDGYVAPLTCSP 1135
Qy 1130 QPEVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVWVDVFAFGAVENPE 1189
Db 1136 QPEVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVWVDVFAFGAVENPE 1195
Qy 1190 YLTPQGGAAQPPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
Db 1196 YLTPQGGAAQPPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
XX
DT 18-DEC-2001 (first entry)
XX
XX Human tyrosine kinase-type receptor, HER-2.
XX
XX Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Region 774..782
FT /note= "Antigenic epitope"
XX
XX WO200168677-A2.
XX
XX PD 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US40328.
XX
XX PR 16-MAR-2000; 2000US-0527487.
XX
XX (GENZ) GENZYME CORP.
XX
XX Nicolette CA;
XX
XX WPI; 2001-616284/71.
DR N-PSDB; AAD19731.
XX
XX Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties
PT

XX
PS
XX
XX Claim 4; Page 63-67; 69pp; English.
XX
CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;
Query Match 97.5%; Score 6625; DB 22; Length 1255;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1222; Conservative 6; Mismatches 16; Indels 16; Gaps 2;
Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPTHLDMLRHLVQCQVVGNL 60
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPTHLDMLRHLVQCQVVGNL 60
Qy 61 ELTYLPTNASLSFLQDQEVQGYVLIHNVQVPLQRLIRVGTQLPEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQDQEVQGYVLIHNVQVPLQRLIRVGTQLPEDNYALAVLNG 120
Qy 121 DPLNNTTPTVGASPGGLRELQRLSITELTKGVLIOIRNPOLCYQDITLWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSITELTKGVLIOIRNPOLCYQDITLWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVAGGACARCKGPIPTDCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVAGGACARCKGPIPTDCHEQC 240
Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGYTTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGYTTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHNVTAEDGTQCEKCKPCARVCYGLGMOYIKANSFEGIT 360
Db 301 YNYLSTDVGSCTLVCPLNHNVTAEDGTQCEKCKPCARVCYGLGMEHLREVAVTSAN 360
Qy 361 ELEPAGCKKIFGSLAFLPESFDGDPASNTAPLQEQVFTLEITGYLYISAMPDSL 420
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQEQVFTLEITGYLYISAMPDSL 420
Qy 421 DLSVFQNLQVIRGRIHLNGAYSILTQGLGISWLGRLSRLRELGSGLALIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRIHLNGAYSILTQGLGISWLGRLSRLRELGSGLALIHNNHLCFVHTV 480
Qy 481 PWDOLFENPHQALLHTANRPEDECVGEGLAHQICARGHCWGPGTQCVCNQSLRGQEC 540
Db 481 PWDOLFENPHQALLHTANRPEDECVGEGLAHQICARGHCWGPGTQCVCNQSLRGQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEFCVARC 600
Qy 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDDDDKGCPCAEQASPLTSIYIKAN 660
Db 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDDDDKGCPCAEQASPLTSIYIKAN 665
Qy 661 SKFIGI-----TELKROOKIRKTYMRLLQETELVEPLTSPGAMPNQAQMRIL 709
Db 656 SAVVGILLVVLGVVFGILIKRROOKIRKTYMRLLQETELVEPLTSPGAMPNQAQMRIL 715


```
Qy 710 KETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMA 769
Db 716 KETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMA 775
Qy 770 GVGSPYVSRLLGICLTSTVQLVTLMPYGCCLLDHVRNRRGLSGQDLNMCQIAKMSY 829
Db 776 GVGSPYVSRLLGICLTSTVQLVTLMPYGCCLLDHVRNRRGLSGQDLNMCQIAKMSY 835
Qy 830 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESIL 889
Db 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESIL 895
Qy 890 RRRFTHQSDVNSYGVTVNLMFTGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYIM 949
Db 896 RRRFTHQSDVNSYGVTVNLMFTGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYIM 955
Qy 950 VKCMWIDSECRPRFRELSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1009
Db 956 VKCMWIDSECRPRFRELSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1015
Qy 1010 DLVDAEEYLVPOQGFCDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSP 1069
Db 1016 DLVDAEEYLVPOQGFCDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSP 1075
Qy 1070 APSEGAGSDVFDGLGMAAGLQSLPTHDPSLQRYSEDPTVPLPSETDGYVAPLTCSP 1129
Db 1076 APSEGAGSDVFDGLGMAAGLQSLPTHDPSLQRYSEDPTVPLPSETDGYVAPLTCSP 1135
Qy 1130 QBEYNQPDVRRPQSPREGPLPAAPAGATLERAKTILSPGKNGVVKDVFAGGAVENPE 1189
Db 1136 QBEYNQPDVRRPQSPREGPLPAAPAGATLERAKTILSPGKNGVVKDVFAGGAVENPE 1195
Qy 1190 YLTPOGGAAPQHPHPAPFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYLGIDVPV 1249
Db 1196 YLTPOGGAAPQHPHPAPFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 3
ID AAB60167 standard; Protein; 1255 AA.
XX
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
OS Homo sapiens.
XX
OS Synthetic.
XX
PN WO200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
PA (GETH ) GENENTECH INC.
XX
PI Erickson S, Schwall R;
XX
DR WPI: 2001-061962/07.
XX
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX
```

PS Example 3; Fig 4; 92pp; English.

CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

SQ Sequence 1255 AA;

Query Match 97.5%; Score 6625; DB 22; Length 1255;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1222; Conservative 6; Mismatches 16; Indels 16; Gaps 2;

```
Qy 1 MELAALCWGGLLALLPPGAASCTGCTDMKRLPASPETHDMLRHLYQCQVVOGNL 60
Db 1 MELAALCWGGLLALLPPGAASCTGCTDMKRLPASPETHDMLRHLYQCQVVOGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQYVLI AHNVQVQLRLRIVRGTQLFEDNYALVDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQYVLI AHNVQVQLRLRIVRGTQLFEDNYALVDNG 120
Qy 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRNPQLCYQDTILWKDIFHKQNQLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRNPQLCYQDTILWKDIFHKQNQLA 180
Qy 181 LTLIDTNSRACHPCSPKSGRSGESSEDCOSLTRTVACGCCARCKGPLTDCCHEOC 240
Db 181 LTLIDTNSRACHPCSPKSGRSGESSEDCOSLTRTVACGCCARCKGPLTDCCHEOC 240
Qy 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 300
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 300
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMQVIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMHLREVRAVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFLPESPDGPASNTAPLOEQLOVFLEETITGLYLSAWPDSL 420
Db 361 IOEFAGCKKIFGSLAFLPESPDGPASNTAPLOEQLOVFLEETITGLYLSAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRSLRSLRSLRSLRSLRSLRSLR 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRSLRSLRSLRSLRSLRSLR 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLOGLPREYVYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKOPPCVARC 600
Db 541 VEECRVLOGLPREYVYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKOPPCVARC 600
Qy 601 PSGVKPDLSTYMPIMKFPDEEGACQPCINCHTSCVDLDDKGCPEAQRASPLTSIQYKAN 660
Db 601 PSGVKPDLSTYMPIMKFPDEEGACQPCINCHTSCVDLDDKGCPEAQRASPLTSIV---- 655
Qy 661 SKFIGI-----TELKRRQOKIRKRYTNRRLLOETELVEPLTPSGAMPNQAQMRIL 709
Db 656 SAVVGILLVVLGVVFGILIKRQOKIRKRYTNRRLLOETELVEPLTPSGAMPNQAQMRIL 715
Qy 710 KETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMA 769
Db 716 KETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMA 775
Qy 770 GVGSPYVSRLLGICLTSTVQLVTLMPYGCCLLDHVRNRRGLSGQDLNMCQIAKMSY 829
Db 776 GVGSPYVSRLLGICLTSTVQLVTLMPYGCCLLDHVRNRRGLSGQDLNMCQIAKMSY 835
Qy 830 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESIL 889
Db 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESIL 895
```

Db 836 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHYHAGGKVPKIMMALESL 895
Qy 890 RRRFTHQSDVWSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIM 949
Db 896 RRRFTHQSDVWSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIM 955
Qy 950 VKCMWIDSECRPRFRELVSERMRDQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDMG 1009
Db 956 VKCMWIDSECRPRFRELVSERMRDQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDMG 1015
Qy 1010 DLVDAEYLVPOOGFFCDDPGAPAGAGVHHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1069
Db 1016 DLVDAEYLVPOOGFFCDDPGAPAGAGVHHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1075
Qy 1070 APSEGAGSDVDFDGLGMAAGKQLSLPHTDPSPLQRYSEDPTVPLPSETDGVYAPLTCSP 1129
Db 1076 APSEGAGSDVDFDGLGMAAGKQLSLPHTDPSPLQRYSEDPTVPLPSETDGVYAPLTCSP 1135
Qy 1130 OPEYVYNQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAGGAVENPE 1189
Db 1136 OPEYVYNQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAGGAVENPE 1195
Qy 1190 YLTPOGGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGDLVPV 1249
Db 1196 YLTPOGGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGDLVPV 1255

RESULT 4

AAU74545
ID AAU74545 standard; Protein; 1255 AA.

AC AAU74545;

XX

XX 23-APR-2002 (first entry)

XX Human HER2 (ErbB2) polypeptide.

XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;
XX anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
XX stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
XX thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
XX glial disorder; astrocytoma; glioma; hypothalamic disorder;
XX glandular disorder; macropapillary disorder; epithelial disorder;
XX stromal disorder; blastocyst disorder; inflammatory disorder;
XX angiogenic disorder; immunological disorder.

XX Homo sapiens.

OS US2002001587-A1.

XX 03-JAN-2002.

XX 16-MAR-2001; 2001US-0811123.

XX 16-MAR-2000; 2000US-189844P.

XX 05-OCT-2000; 2000US-238327P.

XX (ERIC/) ERICKSON S.

XX (SCHW/) SCHWALL R.

XX (SLIW/) SLIWOWSKI M.

XX Erickson S, Schwall R, Sliwowski M;

XX WPI; 2002-163686/21.

XX N-PSDB; ABK14058.

XX Treating tumour characterised by overexpression of epidermal growth

XX factor receptor, ErbB or cancer in mammal, comprises administering

XX anti-ErbB antibody-maytansinoid conjugate to the mammal

XX Example 3; Fig 7; 93pp; English.

XX The invention relates to treating a tumour in a mammal, where the tumour

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytoma, hypothalamic, glandular, macropapillary,
CC epithelial, stromal, blastocyst, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.

XX Sequence 1255 AA;

Query Match 97.5%; Score 6625; DB 23; Length 1255;

Best Local Similarity 97.0%; Pred. No. 0;

Matches 1222; Conservative 6; Mismatches 16; Indels 16; Gaps 2;

Qy 1 MELAALCRWGLLLALLPFGAASVTCTGDMKRLPASPETHLDMRLHLYQGCQVQGNL 60

Db 1 MELAALCRWGLLLALLPFGAASVTCTGDMKRLPASPETHLDMRLHLYQGCQVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVQGVYVLIHNNQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIOEVQGVYVLIHNNQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120

Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLEILKGVLTQRPOLCYQDTILWKDIFHKNNQLA 180

Db 121 DPLNNTTPTVTCASPGGLRELQRLSLEILKGVLTQRPOLCYQDTILWKDIFHKNNQLA 180

Qy 181 LTLDTNRSRACHPCSPCKSGRSGESSEDQSLTRTVACGACARCKGPLTDCCHQEC 240

Db 181 LTLDTNRSRACHPCSPCKSGRSGESSEDQSLTRTVACGACARCKGPLTDCCHQEC 240

Qy 241 AAGTGPXKSHDCLACLFHNSGICELHCPALVYNTDTFESNPNEGRTYFGASCVTACP 300

Db 241 AAGTGPXKSHDCLACLFHNSGICELHCPALVYNTDTFESNPNEGRTYFGASCVTACP 300

Qy 301 YNYLSTVDGSCVLCVPLHNOEVAEDGTQRCCKSPCARVCYGLQMYIKANSKFIGHT 360

Db 301 YNYLSTVDGSCVLCVPLHNOEVAEDGTQRCCKSPCARVCYGLQMYIKANSKFIGHT 360

Qy 361 ELEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVFTEELTYLISAMPDLSLP 420

Db 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVFTEELTYLISAMPDLSLP 420

Qy 421 DLSVFNQLQVIRGRILHNGAYSLLTQGLGI SWLGLRSRLRELGSGLALIHNNTHLCFVHTV 480

Db 421 DLSVFNQLQVIRGRILHNGAYSLLTQGLGI SWLGLRSRLRELGSGLALIHNNTHLCFVHTV 480

Qy 481 PWDOLFNRPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGTQCVCNCSQFIRGQBC 540

Db 481 PWDOLFNRPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGTQCVCNCSQFIRGQBC 540

Qy 541 VEECRVLQGLPREYVYNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPCVARC 600

Db 541 VEECRVLQGLPREYVYNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPCVARC 600

Qy 601 PSGVKPDLSPYMPIWKPFDEEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIQYIKAN 660

Db 601 PSGVKPDLSPYMPIWKPFDEEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIQYIKAN 660

Qy 661 SKFTGI-----TELKROOKIRKYTNRLLOSTELVEPLTPSGAMPNQAMRIL 709

Db 661 SKFTGI-----TELKROOKIRKYTNRLLOSTELVEPLTPSGAMPNQAMRIL 709

Qy 710 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIIPVAIKVLRNTSPKANKELDEAYYMA 769

Db 710 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIIPVAIKVLRNTSPKANKELDEAYYMA 769

Qy 770 GVGSPPYVRLGLICTSTVQLVTLQMPYGCILLDVRNRLGSLGSDLLNMCQIAKGMSY 829

Db 770 GVGSPPYVRLGLICTSTVQLVTLQMPYGCILLDVRNRLGSLGSDLLNMCQIAKGMSY 829

```
Db 776 GVGSPYVSRLLGICLTSTVQLVTLMPYVGLLDHVRNRRGLSGDILLNWCQIAKMSY 835
Qy 830 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 889
Db 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 895
Qy 890 RRRFTHQSDVMSGYVTVMELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIM 949
Db 896 RRRFTHQSDVMSGYVTVMELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIM 955
Qy 950 VKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDMG 1009
Db 956 VKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDMG 1015
Qy 1010 DLVDAEEYLVPOQGFCDPAPGAGGVVHRRSSSTRSGGDLTLGLEPSEEEAPRSP 1069
Db 1016 DLVDAEEYLVPOQGFCDPAPGAGGVVHRRSSSTRSGGDLTLGLEPSEEEAPRSP 1075
Qy 1070 APSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1129
Db 1076 APSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1135
Qy 1130 QPEYVNPDPVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVXDVFAFGGAVENPE 1189
Db 1136 QPEYVNPDPVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVXDVFAFGGAVENPE 1195
Qy 1190 YLTPOGGAAPQHPHPPAFSPAFDNLVYNDODDPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
Db 1196 YLTPOGGAAPQHPHPPAFSPAFDNLVYNDODDPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
```

RESULT 5

```
AAW01111
ID AAW01111 standard; Protein; 1255 AA.
AC AAW01111;
XX
DT 01-JAN-1997 (first entry)
XX
DE HER-2/neu protein.
XX
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS Homo sapiens.
XX
FH Key
FT Domain
FT 676..1255
FT /label= Intracellular domain
FT /note= "claimed domain, useful for immunisation"
XX
PN WO9630514-A1.
XX
PD 03-OCT-1996.
XX
PF 28-MAR-1996; 96WO-US01689.
XX
PR 31-MAR-1995; 95US-0414417.
XX
PA (UNIW ) UNIV WASHINGTON.
XX
PI Cheever MA, Disie ML;
XX
DR WPI; 1996-455361/45.
XX
DR N-PSDB; AAT40739.
XX
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
PS Claim 2; Page 56-61; 71pp; English.
```

```
XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transfected host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 97.4%; Score 6619; DB 17; Length 1255;
Best Local Similarity 96.9%; Pred No. 0;
Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;
Qy 1 MELAALCRWGLLLALLPPGAASVTCTGTDMLKRLPASPETHLDMLRHLVQSCVVGQNL 60
Db 1 MELAALCRWGLLLALLPPGAASVTCTGTDMLKRLPASPETHLDMLRHLVQSCVVGQNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIANQVQVPLQRLRIVRGTLQDPEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIANQVQVPLQRLRIVRGTLQDPEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRPOLCVQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRPOLCVQDTILWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCARGKGPLTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCARGKGPLTDCCHEQC 240
Qy 241 AAGCTGPHGSDCLACLFHNSGICELHCPALVYNTDTFESNPNEGRTFEGASCVTACP 300
Db 241 AAGCTGPHGSDCLACLFHNSGICELHCPALVYNTDTFESNPNEGRTFEGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHNLQEVTAEDGTQRCCKSPCARVCYGLCMQVIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHNLQEVTAEDGTQRCCKSPCARVCYGLCMQHLREVRVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOEQLOVFLEETIETGYLIYSAMPDLSL 420
Db 361 IQEFAAGCKKIFGSLAFPLPESFDGDPASNTAPLOEQLOVFLEETIETGYLIYSAMPDLSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGLSLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGLSLALIHNNTHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECYVGEGLACHOLCARGHCWGPPTQCVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECYVGEGLACHOLCARGHCWGPPTQCVCNCSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFFCVARC 600
Db 541 VEECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFFCVARC 600
Qy 601 PSGVKPDLISYMPIWKFPDEEGACQCPINCHTSCVDLDDKGPAPORASPLTSIQIYKAN 660
Db 601 PSGVKPDLISYMPIWKFPDEEGACQCPINCHTSCVDLDDKGPAPORASPLTSI-----I 655
Qy 661 SKFIGI-----TELKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRIL 709
Db 656 SAVGILLVVLGVVFGVILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRIL 715
Qy 710 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDEAYVMA 769
Db 716 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDEAYVMA 775
Qy 770 GVGSPYVSRLLGICLTSTVQLVTLMPYVGLLDHVRNRRGLSGDILLNWCQIAKMSY 829
Db 776 GVGSPYVSRLLGICLTSTVQLVTLMPYVGLLDHVRNRRGLSGDILLNWCQIAKMSY 835
```

QY 830 LEDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 889
 DB 836 LEDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 895
 QY 890 RRRFTHQSDVMSYGVTVNHELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIM 949
 DB 896 RRRFTHQSDVMSYGVTVNHELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIM 955
 QY 950 VKCWMIDSECRPRFRELVSERMDARDQRFVVIQNEDLGPASPLDSTFYRSLLEDMDG 1009
 DB 956 VKCWMIDSECRPRFRELVSERMDARDQRFVVIQNEDLGPASPLDSTFYRSLLEDMDG 1015
 QY 1010 DLVDAEYLVPOQGFCDPAPGAGMHHRRSSSTRSGGDLTLGLEPSEERAPRSP 1069
 DB 1016 DLVDAEYLVPOQGFCDPAPGAGMHHRRSSSTRSGGDLTLGLEPSEERAPRSP 1075
 QY 1070 APSEGAGSDVFDGDLGMAAGLQSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCS 1129
 DB 1076 APSEGAGSDVFDGDLGMAAGLQSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCS 1135
 QY 1130 QPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPE 1189
 DB 1136 QPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPE 1195
 QY 1190 YLTPQCGAAPPHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGPTAENPEYLGLDV 1249
 DB 1196 YLTPQCGAAPPHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGPTAENPEYLGLDV 1255

RESULT 6
 AA92406
 ID AA92406 standard; Protein; 1255 AA.
 XX AA92406;
 AC AA92406;
 XX AA92406;
 DT 21-APR-1999 (first entry)
 XX Human HER-2/neu oncogene protein.
 DE HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 KW malignancy; treatment; tumour.
 KW Homo sapiens.
 OS Homo sapiens.
 PH Key Location/Qualifiers
 FT Region 676..1255
 FT /note= "region which elicits immune response"
 XX US5869445-A.
 XX 09-FEB-1999.
 XX 01-APR-1996; 96US-0625101.
 XX 01-APR-1996; 96US-0625101.
 PR 17-MAR-1993; 93US-0033644.
 PR 12-AUG-1993; 93US-0106112.
 PR 31-MAR-1995; 95US-0414417.
 XX (UNIW) UNIV WASHINGTON.
 XX Cheever MA, Disis ML;
 XX WPT; 1999-152835/13.
 DR N-PSDB; AAX01912.
 XX Use of HER-2/neu polypeptides - for eliciting an immune response to
 PT an HER-2/neu associated malignancy, particularly for treating or
 PT preventing tumours
 XX Claim 3; Column 31-38; 26pp; English.
 PS This sequence represents the human HER-2/neu oncogene protein. A fragment
 XX

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or reoccurrence.
 XX

SQ Sequence 1255 AA;

Query Match 97.4%; Score 6619; DB 20; Length 1255;
 Best Local Similarity 96.9%; Pred. No. 0;
 Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;

QY 1 MELAALCRWGLLLALLPFGAASQTCTGDMKRLRLPASPETHDLMLRLHYQCCVQVQNL 60
 DB 1 MELAALCRWGLLLALLPFGAASQTCTGDMKRLRLPASPETHDLMLRLHYQCCVQVQNL 60
 QY 61 ELTYLPTNASLSFLQDIQEVQYVLIHANOVQVPLQRLIRIVRGTLQEDNYALAVLDNG 120
 DB 61 ELTYLPTNASLSFLQDIQEVQYVLIHANOVQVPLQRLIRIVRGTLQEDNYALAVLDNG 120
 QY 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
 DB 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
 QY 181 LTLIDTNRSRACHPCSPMKGSRMGESSEDCQSLTRTVAGGCARCKGPLTDCCHQEC 240
 DB 181 LTLIDTNRSRACHPCSPMKGSRMGESSEDCQSLTRTVAGGCARCKGPLTDCCHQEC 240
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMQYIKANSKFEGIT 360
 DB 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMHELRVAVTSAN 360
 QY 361 ELEFAGCKKIFGSLAFPESEFDGPASNTAPLQEPQLQVFTLEBITGYLYISAMPDSLP 420
 DB 361 IQEFAGCKKIFGSLAFPESEFDGPASNTAPLQEPQLQVFTLEBITGYLYISAMPDSLP 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLSLQGLGISWGLRLSRLRELGSLALHNNHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSLSLQGLGISWGLRLSRLRELGSLALHNNHLCFVHTV 480
 QY 481 PWDQLFRNPHOALLHTANRDECEVGEGLACHQLCARGHCWGPGTQCVCNCSQFLRGQSC 540
 DB 481 PWDQLFRNPHOALLHTANRDECEVGEGLACHQLCARGHCWGPGTQCVCNCSQFLRGQSC 540
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEFCVARC 600
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEFCVARC 600
 QY 601 PSGVKPDLSPYMPIWKFPDEEGACPCPINCTHSCVDDDDKGCAPAQASPLTSIQYIKAN 660
 DB 601 PSGVKPDLSPYMPIWKFPDEEGACPCPINCTHSCVDDDDKGCAPAQASPLTSI-----I 655
 QY 661 SKFIGI-----TELKROOKIRKYTMRLLOETELVEPLTPSGAMPNQMRIL 709
 DB 656 SAVVGIILVVVLGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAMPNQMRIL 715
 QY 710 KETELRKVKVLGSGAFGTVYKGIWIPDENVKIPIVAIKVLRNTPSKANKETLDEAYMA 769
 DB 716 KETELRKVKVLGSGAFGTVYKGIWIPDENVKIPIVAIKVLRNTPSKANKETLDEAYMA 775
 QY 770 GVGSPPYVRLIGICLTSTVQLVTQMPYVGLLDHVRNRLGSLQDGLNWCQIAKMSY 829
 DB 776 GVGSPPYVRLIGICLTSTVQLVTQMPYVGLLDHVRNRLGSLQDGLNWCQIAKMSY 835
 QY 830 LEDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 889
 DB 836 LEDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 895

Db	956	VKCHMIDSECPREPRFELVSEFRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDDDMG	1015
Qy	1010	DLVDAEYLVPOQGFCCPDPAAGAGGMVHHRHSSSTRSGGDLTLGLEPSEEEAPRSPL	1069
Db	1016	DLVDAEYLVPOQGFCCPDPAAGAGGMVHHRHSSSTRSGGDLTLGLEPSEEEAPRSPL	1075
Qy	1070	APSEGAGSDVFDGDLGNGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP	1129
Db	1076	APSEGAGSDVFDGDLGNGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP	1135
Qy	1130	QPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTSLSPKNGVVKDVPAPGGAVENTPE	1189
Db	1136	QPEYVNPQDVRPQPPSPREGPLPAARPAGATLERKPTLSPEKNGVVKDVPAPGGAVENTPE	1195
Qy	1190	YLTPQGAAQPQHPPPAFSPAFONLYYWDQDPPERGAPPSTFKGTPTAENPEYVLGLDVPV	1249
Db	1196	YLTPQGAAQPQHPPPAFSPAFONLYYWDQDPPERGAPPSTFKGTPTAENPEYVLGLDVPV	1255

RESULT 8
 AAY84780
 ID AAY84780 standard; Protein; 1255 AA.

Query Match	97.4%	Score 6619;	DB 21;	Length 1255;
Beat Local Similarity	96.9%	Mismatches 6;	Mismatches 17;	Indels 16;
Matches 1221;	Conservative	6;	Indels	Gaps 2;
Qy	1	MELAAALCRWGLLLALLPPGAAS	TQVCTGTDMKRLPASPETHLDMRLHYOGCQVQGNL	60
Db	1	MELAAALCRWGLLLALLPPGAAS	TQVCTGTDMKRLPASPETHLDMRLHYOGCQVQGNL	60
Qy	61	ELTYLPTNASLSFLQDIOEQVGY	LIAHNQVRQVPLQRLRIVRG	120
Db	61	ELTYLPTNASLSFLQDIOEQVGY	LIAHNQVRQVPLQRLRIVRG	120
Qy	121	DPLNNTTPVTGASPGGLREQLR	LSLTELKGGVLIQRNPOLCYQD	180
Db	121	DPLNNTTPVTGASPGGLREQLR	LSLTELKGGVLIQRNPOLCYQD	180
Qy	181	LTLTDNRSRACHPCSPMKGSR	CMWSSSDCQSLTRTVCAGGCAR	240
Db	181	LTLTDNRSRACHPCSPMKGSR	CMWSSSDCQSLTRTVCAGGCAR	240
Qy	241	AAGCTGPKHSDCLACHFNHSGI	CELHCPALVTYNTDTFESMPN	300
Db	241	AAGCTGPKHSDCLACHFNHSGI	CELHCPALVTYNTDTFESMPN	300
Qy	301	YNYLSTDVGSCTLVCPHLNHQ	EVTAEDGTQCEKSKPCARVCY	360
Db	301	YNYLSTDVGSCTLVCPHLNHQ	EVTAEDGTQCEKSKPCARVCY	360
Qy	361	ELEFAGCKKIFGSLAFIPESFD	GDPASNTAPLQPEQLQVETL	420
Db	361	IQEFAGCKKIFGSLAFIPESFD	GDPASNTAPLQPEQLQVETL	420
Qy	421	DLVSFQNLQVIRGRILHNGAYS	LTLOGLGTSLWGLRSLRELGS	480
Db	421	DLVSFQNLQVIRGRILHNGAYS	LTLOGLGTSLWGLRSLRELGS	480
Qy	481	PWDOLFNRPHQALLHTANR	PEDEBCVGEGLACHQLCARG	540
Db	481	PWDOLFNRPHQALLHTANR	PEDEBCVGEGLACHQLCARG	540
Qy	541	VEECRVLQGLPREVYNARHCL	PCHEPCOPONGSVTCGPEAD	600
Db	541	VEECRVLQGLPREVYNARHCL	PCHEPCOPONGSVTCGPEAD	600
Qy	601	PSGVKPDLSYMPKPFDEBGA	QCPCPINCTHSCVDLDDK	660
Db	601	PSGVKPDLSYMPKPFDEBGA	QCPCPINCTHSCVDLDDK	660
Qy	661	SKFTGI-----TELKRRQ	QKIRKYMTRLLQETELVE	709
Db	656	SAVVGILLVVVLGVVFGILIK	RQKIRKYMTRLLQETELVE	715
Qy	710	KETELRKVKVLGSGARGTVY	KGWIIPDGENVKIPVAIKV	769
Db	716	KETELRKVKVLGSGARGTVY	KGWIIPDGENVKIPVAIKV	775
Qy	770	GVGSPVYSRLIGLCLSTVOL	VTQLMPYGCLLDHHVNRN	829
Db	776	GVGSPVYSRLIGLCLSTVOL	VTQLMPYGCLLDHHVNRN	835
Qy	830	LEDVRLVHRDLAARNVLVKS	PNHVKITDFGLARLLDIT	889
Db	836	LEDVRLVHRDLAARNVLVKS	PNHVKITDFGLARLLDIT	895
Qy	890	RRRFTHQSDVWSYGVTVW	ELMTGAKPYDGIIPAREI	949
Db	896	RRRFTHQSDVWSYGVTVW	ELMTGAKPYDGIIPAREI	955
Qy	950	VKCMWMLDSECRPFREL	VSEFSRMARDPQRFVVI	1009
Db	956	VKCMWMLDSECRPFREL	VSEFSRMARDPQRFVVI	1015
Qy	1010	DLVDAEYLVPQOGFCFDP	PAPGAGGMVHRHRSST	1069

Db 1016 DLVDAEYLVPOQFFCPDPAPGAGMVRHRRSSSTRSGGDLTLGLEPSEEA PRSP 1075
Qy 1070 APSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLT CSP 1129
Db 1076 APSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLT CSP 1135
Qy 1130 QPEYVYQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFARFGGAVENPE 1189
Db 1136 QPEYVYQPDVVRPQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVVKDVFARFGGAVENPE 1195
Qy 1190 YLTPOGGAAPQHPHPPAPSPAFDNLYYWDQDPPERGAPSTPKGTPTAENPEYLGLDV PV 1249
Db 1196 YLTPOGGAAPQHPHPPAPSPAFDNLYYWDQDPPERGAPSTPKGTPTAENPEYLGLDV PV 1255

RESULT 9
ID AAB85458 standard; Protein; 1255 AA.
XX AAB85458;
XX 25-SEP-2001 (first entry)
XX Human HER-2/neu protein.
DE Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX Homo sapiens.
XX WO200153463-A2.
XX 26-JUL-2001.
XX 19-JAN-2001; 2001WO-US01850.
XX 21-JAN-2000; 2000US-0177545.
XX (COR1-) CORIXA CORP.
XX Cheever MA, Hand-Zimmermann S;
XX WPI; 2001-476112/51.
XX N-PSDB; AAH23392.
XX New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer -
XX Claim 2; Page 41-46; 49pp; English.
XX The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX Sequence 1255 AA;

Query Match 97.4%; Score 6619; DB 22; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;
Qy 1 MELAAALCRWGLLLALLPPGAASTQVCTGTDMLKRLPASPTHLDMLRHLHYQCQCVVQGNL 60
Db 1 MELAAALCRWGLLLALLPPGAASTQVCTGTDMLKRLPASPTHLDMLRHLHYQCQCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQLRLIRVRGTQLPEDNVALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQLRLIRVRGTQLPEDNVALAVLDNG 120
Qy 121 DPLNNTTVPVTGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIPHKNNQLA 180
Db 121 DPLNNTTVPVTGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIPHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSRCHGSESDCOSLTRTVCAGGCARCKGKPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSRCHGSESDCOSLTRTVCAGGCARCKGKPLPTDCCHEQC 240
Qy 241 AAGTGPXHSBCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGTGPXHSBCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMOVIKANSRFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEETGYLYISAWPDSLP 420
Db 361 IOEFAGCKKI FGS LAFLPESFDGDPASNTAPLOPEQLQVFETLEETGYLYISAWPDSLP 420
Qy 421 DLSVPQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLGALIHNNTHLCFVHTV 480
Db 421 DLSVPQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLGALIHNNTHLCFVHTV 480
Qy 481 PMDQLFRNPHQALLHTANRPEDECYEGGLACHOLCARGHCWGPGTQCNCQFLRGQBC 540
Db 481 PMDQLFRNPHQALLHTANRPEDECYEGGLACHOLCARGHCWGPGTQCNCQFLRGQBC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCACAHYKDPFFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCACAHYKDPFFCVARC 600
Qy 601 PSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGPAPQORASPLTSIQYIKAN 660
Db 601 PSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGPAPQORASPLTSI-----I 655
Qy 661 SKFIGI-----TELKROQKIRKYMRRLLQETELVEPLTPSGAMNQAMRIL 709
Db 656 SAVVGILLVWVGVLGKIRKYMRRLLQETELVEPLTPSGAMNQAMRIL 715
Qy 710 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKILDEAYVMA 769
Db 716 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKILDEAYVMA 775
Qy 770 GVGSPYVSRLLGICLTSTVQLVQVLTQVLMGYCGLLDHVRENRLGSLGQDLNWCQIAKMSY 829
Db 776 GVGSPYVSRLLGICLTSTVQLVQVLTQVLMGYCGLLDHVRENRLGSLGQDLNWCQIAKMSY 835
Qy 830 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLDDIDETEHADGGKVPKIMWALESTIL 889
Db 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLDDIDETEHADGGKVPKIMWALESTIL 895
Qy 890 RRRFTHQSDVMSGYVTVMELMTFGAKPYDGIIPAREIPOLLEKGERLPQPICTIDVYIM 949
Db 896 RRRFTHQSDVMSGYVTVMELMTFGAKPYDGIIPAREIPOLLEKGERLPQPICTIDVYIM 955
Qy 950 VKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1009
Db 956 VKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1015
Qy 1010 DLVDAEYLVPOQFFCPDPAPGAGMVRHRRSSSTRSGGDLTLGLEPSEEA PRSP 1069
Db 1016 DLVDAEYLVPOQFFCPDPAPGAGMVRHRRSSSTRSGGDLTLGLEPSEEA PRSP 1075
Qy 1070 APSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLT CSP 1129
Db 1076 APSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLT CSP 1135
Qy 1130 QPEYVYQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFARFGGAVENPE 1189

Db 1136 QPEYVNPQDVRFQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVVKVFAFGGAVENPE 1195
 QY 1190 YLTPOGGAPOHPHPPAFSPADNLYYWDQDPPERCAPSTFKGTPTAENPEYLGJDPV 1249
 Db 1196 YLTPOGGAPOHPHPPAFSPADNLYYWDQDPPERCAPSTFKGTPTAENPEYLGJDPV 1255

RESULT 10

AAG88267

ID AAG88267 standard; Protein; 1255 AA.

XX

AC AAG88267;

DT 11-SEP-2001 (first entry)

XX

DE HER2/new amino acid sequence.

XX

KW Human; HER2/new; epitope; human leukocyte antigen; HLA; T cell;

KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;

KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.

XX

OS Homo sapiens.

XX

PN WO200141787-A1.

PD 14-JUN-2001.

XX

PF 11-DEC-2000; 2000WO-US33591.

XX

PR 10-DEC-1999; 99US-0458299.

XX

PA (EPIM-) EPIMUNE INC.

XX

PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Cellis E;

PI Keogh E;

XX

DR WPI; 2001-374995/39.

XX

PT An isolated prepared HER2/new epitope useful in a vaccine for inducing

PT cellular immune responses for the prevention and treatment of cancer -

XX

PS Disclosure; Page 15; 199pp; English.

XX

The present invention describes isolated prepared HER2/new epitopes (I). Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is culture in vitro and binds to a complex of an epitope (I), bound to a human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I) and a second epitope and the peptide is less than 50 contiguous amino acids that have 100% identity with a native peptide sequence of HER2/new; (3) a vaccine composition (III) comprising (II) and a pharmaceutical excipient; (4) an isolated nucleic acid encoding a peptide comprising (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and immunostimulant activities, and can be used in vaccines. (I), (II) and (III) are useful for inducing cellular immune responses for the prevention and treatment of cancer. (I) and (II) are useful for monitoring or evaluating an immune response to a tumour-associated antigen when incubated with a T lymphocyte sample form a patient and detecting the presence of bound T lymphocyte to (I) or (II). Epitope based vaccines mean that immunosuppressive epitopes that may be present in whole antigens may be avoided. Selected epitopes may be combined to enhance immunogenicity. The possible pathological side effects caused by infectious agents or whole protein antigen is eliminated. The vaccine provides the ability to direct and focus an immune response to multiple selected antigens from the same pathogen. Epitope-based anti-tumour vaccines provides the opportunity to combine epitopes derived from multiple tumour-associated molecules addressing the problem of tumour tumour variability and reducing the likelihood of tumour escape due to antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in the exemplification of the present invention.

SQ Sequence 1255 AA;

Query Match

97.4%; Score 6619; DB 22; Length 1255;

Best Local Similarity 96.9%; Pred. No. 0;
 Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;
 QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPEHLDMLRHLHQGQVVOGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPEHLDMLRHLHQGQVVOGNL 60
 QY 61 ELTYLPTNASLSFLODIQEVQGVLIHNNQVQVPLQRLIRVRGTOLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLODIQEVQGVLIHNNQVQVPLQRLIRVRGTOLFEDNYALAVLDNG 120
 QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFHKNOLA 180
 Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFHKNOLA 180
 QY 181 LTLIDTNRSRACHPCSPCKGSRGSESSDCQSLTRTVTCAGGCARCKGKPLPTDCCHEQC 240
 Db 181 LTLIDTNRSRACHPCSPCKGSRGSESSDCQSLTRTVTCAGGCARCKGKPLPTDCCHEQC 240
 QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPHNNQEVTAEDGTORCEKSKPCARVCYGLGQVIKANSKFIGIT 360
 Db 301 YNYLSTDVGSCTLVCPHNNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
 QY 361 ELEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPELOQVFETLEETIGVLYISAWPDSL 420
 Db 361 IQEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPELOQVFETLEETIGVLYISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLISWGLRSLRELGLSLALIHNNTHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLISWGLRSLRELGLSLALIHNNTHLCFVHTV 480
 QY 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVCNCSOFLRGQEC 540
 Db 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVCNCSOFLRGQEC 540
 QY 541 VEECRVLQGLPREYVYNARHCLPCHPECPQNGSVTCFGEADQCVCACAHYKDPFFCVARC 600
 Db 541 VEECRVLQGLPREYVYNARHCLPCHPECPQNGSVTCFGEADQCVCACAHYKDPFFCVARC 600
 QY 601 PSGVKPDLISYMPIWKPDEEGACQPCPINCTHSCVDLDDKGCPCAEORASPLTSIQYKAN 660
 Db 601 PSGVKPDLISYMPIWKPDEEGACQPCPINCTHSCVDLDDKGCPCAEORASPLTSI ----I 655
 QY 661 SKFIGI-----TELKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRIL 709
 Db 661 SKFIGI-----TELKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRIL 715
 QY 710 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELDEAYVMA 769
 Db 716 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELDEAYVMA 775
 QY 770 GVGSYVSRLLGICLTSTVQLVTLMPYGCCLLDHVRNRRGLSGQDLNWCQIAGMSY 829
 Db 776 GVGSYVSRLLGICLTSTVQLVTLMPYGCCLLDHVRNRRGLSGQDLNWCQIAGMSY 835
 QY 830 LEDVRLVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESL 889
 Db 836 LEDVRLVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESL 895
 QY 890 RRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIM 949
 Db 896 RRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIM 955
 QY 950 VKCWMIDSECRPRELVSFESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMG 1009
 Db 956 VKCWMIDSECRPRELVSFESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMG 1015
 QY 1010 DLVDAEEVLVPOQGFPCDDPAPAGGVHHRSSSTRSGGDLTLGLEPSEEEAPRPL 1069

Db 1016 DLVDAEYLVPOQGFCDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSP 1075
QY 1070 ARSEGAGSDVFDGLGMAKGLQSLPTHDPSPLOYSDDPTVLPSETDGVVAPLTCSP 1129
Db 1076 ARSEGAGSDVFDGLGMAKGLQSLPTHDPSPLOYSDDPTVLPSETDGVVAPLTCSP 1135
QY 1130 QPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVVKOVFAFGGAVENPE 1189
Db 1136 QPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVVVKOVFAFGGAVENPE 1195
QY 1190 YLTPOGGAAPQHPHPPAPSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYGLDVVP 1249
Db 1196 YLTPOGGAAPQHPHPPAPSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYGLDVVP 1255

RESULT 11

AAE24067
ID AAE24067 standard; Protein; 1255 AA.

XX AC

AAE24067;

DT 23-SEP-2002 (first entry)

XX DE Human Her-2 protein.

XX KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.

XX OS Homo sapiens.

XX PN WO200222636-A1.

XX PD 21-MAR-2002.

PF 12-SEP-2001; 2001WO-US28572.

PR 15-SEP-2000; 2000US-0663834.

XX PA (ISIS-) ISIS PHARM INC.

XX PI Bennett CF, Cowseert LM;

XX DR N-PSDB; AAD38904.

XX PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -

XX PS Example 13; Page 95-107; 116pp; English.

XX CC The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.

XX SQ Sequence 1255 AA;

Query Match 97.4%; Score 6619; DB 23; Length 1255;

Best Local Similarity 96.9%; Pred. No. 0;

Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLLPASPEHLDMLRHLVQGCQVVGNL 60

Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLLPASPEHLDMLRHLVQGCQVVGNL 60

QY 61 ELTYLPTNASLFLQDIQEVQYVLI AHNQVRQVPLQRLRI VRGTQLFEDNTALVALVDNG 120
Db 61 ELTYLPTNASLFLQDIQEVQYVLI AHNQVRQVPLQRLRI VRGTQLFEDNTALVALVDNG 120
QY 121 DPLNTTPTVGTASPGGLRELQRLSLTEILKGGVLIORNPOLCYODTILWKDIFHKQNOLA 180
Db 121 DPLNTTPTVGTASPGGLRELQRLSLTEILKGGVLIORNPOLCYODTILWKDIFHKQNOLA 180
QY 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTV CAGGCARCKGKPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTV CAGGCARCKGKPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRVTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRVTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAE DGTQRCCKSPCARVCYGLGMQVLIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAE DGTQRCCKSPCARVCYGLGMQVLIKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPELOLVFETLEEITGYLYISAWPDSLP 420
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPELOLVFETLEEITGYLYISAWPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRLSRLSGSLALIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRLSRLSGSLALIHNNHLCFVHTV 480
QY 481 PWDQLFRPHQALLHTANRPEDECYVGEGLACHQLCARGHCWGPPTQVCNCSQFLRGQEC 540
Db 481 PWDQLFRPHQALLHTANRPEDECYVGEGLACHQLCARGHCWGPPTQVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVYNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFFCVARC 600
Db 541 VEECRVLQGLPREYVYNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFFCVARC 600
QY 601 PSGVKPDLSYMPINWKPDEEGACQPCINCTHSCVDLDDKGCPAEQRASPLTSIQIYKAN 660
Db 601 PSGVKPDLSYMPINWKPDEEGACQPCINCTHSCVDLDDKGCPAEQRASPLTSI-----I 655
QY 661 SKFIGI-----TELKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAOMRIL 709
Db 656 SAVVGILLVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAOMRIL 715
QY 710 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMA 769
Db 716 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMA 775
QY 770 GYGSPVSRLLGICLTSTVQLVTLMPYGCCLLDHVRENRRGLSGQDLLNWCQIAKGMSY 829
Db 776 GYGSPVSRLLGICLTSTVQLVTLMPYGCCLLDHVRENRRGLSGQDLLNWCQIAKGMSY 835
QY 830 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 889
Db 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 895
QY 890 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMIM 949
Db 896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMIM 955
QY 950 VKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1009
Db 956 VKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1015
QY 1010 DLVDAEYLVPOQGFCDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSP 1069
Db 1016 DLVDAEYLVPOQGFCDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSP 1075
QY 1070 APSEGAGSDVFDGLGMAKGLQSLPTHDPSPLOYSDDPTVLPSETDGVVAPLTCSP 1129
Db 1076 APSEGAGSDVFDGLGMAKGLQSLPTHDPSPLOYSDDPTVLPSETDGVVAPLTCSP 1135
QY 1130 QPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVVKOVFAFGGAVENPE 1189

Db :136 QPEYVQPDVRPQPSPREGPLAARPAAGATLERPKTUSPGKNGVVKDVFAGGAVENPE 1195
QY 1190 YLTPOGGAAPQHPHPPAFSPAFDNLNYYDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1249
Db 1196 YLTPOGGAAPQHPHPPAFSPAFDNLNYYDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 12
ID AAE20479 standard; Protein; 1255 AA.
AC AAE20479;
XX
DT 01-JUL-2002 (first entry)
XX Human Her-2/neu protein.
DE Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
XX
PN WO200214503-A2.
XX
PD 21-FEB-2002.
XX
PF 14-AUG-2001; 2001WO-US41733.
XX
PR 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-236428P.
PR 21-FEB-2001; 2001US-270520P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedvick TS;
XX
WPI: 2002-280758/32.
DR N-PSDB; AAD32743.
XX
PT Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer -
XX
PS Disclosure; Page 114-117; 129pp; English.
XX
CC The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
XX present sequence is human Her-2/neu protein.
SQ Sequence 1255 AA;

Query Match 97.4%; Score 6619; DB 23; Length 1255;

Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;
QY 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHLYQCQCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHLYQCQCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIHAHQVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIHAHQVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKQNOLA 180
Db 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKQNOLA 180
QY 181 LTLIDTNRSRACHPCSPMKGSRCHGESSEDCQSUTRTVCAGGCARCKGPLPTDCHEOC 240
Db 181 LTLIDTNRSRACHPCSPMKGSRCHGESSEDCQSUTRTVCAGGCARCKGPLPTDCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQCEKSKPCARVCYGLMQYIKANSKFEIGIT 360
Db 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQCEKSKPCARVCYGLMQYIKANSKFEIGIT 360
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFEETLEITGYLYISAMPDLSUP 420
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFEETLEITGYLYISAMPDLSUP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLGISWLGRLSRLRELGSGLALIHNTHLFCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLGISWLGRLSRLRELGSGLALIHNTHLFCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGTQCVCNCSQFLRGQBC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGTQCVCNCSQFLRGQBC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPPECVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPPECVARC 600
QY 601 PSGVKPDLSPYMPKPFDEEGACQPCINCTHSCVDLDDKGCAPQASPLTSIQYIKAN 660
Db 601 PSGVKPDLSPYMPKPFDEEGACQPCINCTHSCVDLDDKGCAPQASPLTSI-----I 655
QY 661 SKFIGI-----TELXRRQOKIRKYTMRRLLOETELVEPLTPSGAMPNQAMRIL 709
Db 656 SAVVGILLVVVVLGVVVGILIKRRQOKIRKYTMRRLLOETELVEPLTPSGAMPNQAMRIL 715
QY 710 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMA 769
Db 716 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMA 775
QY 770 GVGSPYVSRLLGICLTSTVQLVLTQMPYGCILLDHVRENRLGSDLLNWCNQIAKMSY 829
Db 776 GVGSPYVSRLLGICLTSTVQLVLTQMPYGCILLDHVRENRLGSDLLNWCNQIAKMSY 835
QY 830 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESTL 889
Db 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESTL 895
QY 890 RRRRTHOSDVMSYGVTVWELMTFCAPYDGTIPAREIPDLLEKGERLPPOPTCTIDVYIM 949
Db 896 RRRRTHOSDVMSYGVTVWELMTFCAPYDGTIPAREIPDLLEKGERLPPOPTCTIDVYIM 955
QY 950 VKCMWIDSECRPRELVSFSESRMARDPQRVWVQNEDLGPASPLDSTFYFSLLEDDMG 1009
Db 956 VKCMWIDSECRPRELVSFSESRMARDPQRVWVQNEDLGPASPLDSTFYFSLLEDDMG 1015
QY 1010 DLVDAEYLVPOQGFPCDDPAPAGQVMVHHRSSSTRSGGDLTLGLPESEEPASPL 1069

Db 1016 DLVDAEYLVQGGFFCDDPAPAGAGMWHRRSSSTRSGGDLTLGLEPSEEPAPRSPL 1075
 Qy 1070 APSEGAGSDVFDGDLGMAAGLQSLPHDPSPLQRYSEDTVPPLPSETDGVAPLTCSP 1129
 Db 1076 APSEGAGSDVFDGDLGMAAGLQSLPHDPSPLQRYSEDTVPPLPSETDGVAPLTCSP 1135
 Qy 1130 QPEYVNPQDVNRQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVDVFAFGGAVENPE 1189
 Db 1136 QPEYVNPQDVNRQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVDVFAFGGAVENPE 1195
 Qy 1190 YLTQGGGAOPHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYILGLDVPV 1249
 Db 1196 YLTQGGGAOPHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYILGLDVPV 1255

RESULT 13

AAM51143

ID AAM51143 standard; Protein; 1255 AA.

XX AC

XX AAM51143;

XX DT 17-JUN-2002 (first entry)

XX DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
 KW tyrosine kinase; receptor; c-erbB2; gene therapy.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Domain 1..653

XX FT /note= "extracellular domain"

XX FT Domain 676..1255

XX FT /note= "intracellular domain"

XX FT Domain 990..1255

XX FT /note= "phosphorylation domain"

XX XX

XX WO200212341-A2.

XX PN

XX 14-FEB-2002.

XX PD

XX 03-AUG-2001; 2001WO-US24283.

XX PF

XX 03-AUG-2000; 2000US-0632507.

XX PR

XX (CORI-) CORIXA CORP.

XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX XX

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX DR N-PSDB; ABA92250.

XX XX

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
 PT or enhancing an immune response to the protein, has Her-2/neu
 PT extracellular domain fused to Her-2/neu intracellular or
 PT phosphorylation domain -
 XX
 PS Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
 CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
 CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
 CC in a variety of cancers, including breast, ovarian, colon, lung and
 CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
 CC family of receptor-like glycoproteins. It comprises an extracellular
 CC domain with homology to the epidermal growth factor receptor
 CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
 CC intracellular domain that also shows homology to EGFR. Its
 CC overexpression correlates with a poor prognosis in breast and
 CC ovarian cancers. The invention provides Her-2/neu fusion
 CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
 CC preferred fusion proteins, the extracellular domain of a Her-2/neu
 CC protein is fused to a Her-2/neu intracellular domain or
 CC phosphorylation domain (or its DeltapD fragment). An immune
 CC response to Her-2/neu protein is elicited or enhanced by
 CC administering the fusion protein in the form of a vaccine, or by
 CC transfecting cells of an animal ex vivo with a nucleic acid
 CC encoding the fusion protein, and delivering the transfected cells
 CC to the animal. The fusion proteins, nucleic acids, and isolated
 CC specific T-cells are useful for inhibiting the development of a
 CC cancer, especially breast, ovarian, colon, lung or prostate cancer
 CC in a patient. T cells that specifically react with a Her-2/neu
 CC fusion protein can be used to remove tumour cells from a sample in
 CC order to inhibit the development of cancer in a patient.

SQ Sequence 1255 AA;

Query Match 97.4%; Score 6619; DB 23; Length 1255;
 Best Local Similarity 96.9%; Pred. No. 0;
 Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLPASPETHLDMRLHYQSCVQVGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLPASPETHLDMRLHYQSCVQVGNL 60
 Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTLQFEDNYALVDNG 120
 Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRIVRGTLQFEDNYALVDNG 120
 Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHKQNQLA 180
 Db 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHKQNQLA 180
 Qy 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARGKGLPTDCHEQC 240
 Db 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARGKGLPTDCHEQC 240
 Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMQVIKANSFIGIT 360
 Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMHLEVRVTSAN 360
 Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEETGVLYSAMPDSLP 420
 Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEETGVLYSAMPDSLP 420
 Qy 421 DLSVFQNLQVIRGRIHNGAYSLTLQGLGISWLGRLSRLGSLGIALIHHNTHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRIHNGAYSLTLQGLGISWLGRLSRLGSLGIALIHHNTHLCFVHTV 480
 Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVCNCSQFLRGQBC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVCNCSQFLRGQBC 540
 Qy 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFCPEADQCVACAHYKDPPECVARC 600
 Db 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFCPEADQCVACAHYKDPPECVARC 600
 Qy 601 PSGVKPDLSTYMP1WKFPDEEGACQPCPINCTHSCVDLDDKGCAPAQRASPLTSIQVIKAN 660
 Db 601 PSGVKPDLSTYMP1WKFPDEEGACQPCPINCTHSCVDLDDKGCAPAQRASPLTSI 655
 Qy 661 SKFTGI-----TELKROOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRIL 709
 Db 656 SAVVGILLVWVGVVFGILIKRROOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRIL 715
 Qy 710 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYYMA 769
 Db 716 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYYMA 775

QY 770 GVGSPPVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGLSGQDLLNWCQIAKMSY 829
 DB 776 GVGSPPVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGLSGQDLLNWCQIAKMSY 835
 QY 830 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPIKWMALLESIL 889
 DB 836 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPIKWMALLESIL 895
 QY 890 RRRFTHQSDVMSYGVTVNMLMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIM 949
 DB 896 RRRFTHQSDVMSYGVTVNMLMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIM 955
 QY 950 VKCMNIDSECRPRFRELVSFESRMARDPQRFVVIQNEIDLGRASPLDSTFYRSLLEDDMG 1009
 DB 956 VKCMNIDSECRPRFRELVSFESRMARDPQRFVVIQNEIDLGRASPLDSTFYRSLLEDDMG 1015
 QY 1010 DLVDAEYLVPOQGFCCPDAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1069
 DB 1016 DLVDAEYLVPOQGFCCPDAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1075
 QY 1070 APSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSP 1129
 DB 1076 APSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSP 1135
 QY 1130 QPEYVNOPDVROPSPREGPLPAARPAGATLERAKTLSPGKNGVYKDVFAFGGAVENTPE 1189
 DB 1136 QPEYVNOPDVROPSPREGPLPAARPAGATLERAKTLSPGKNGVYKDVFAFGGAVENTPE 1195
 QY 1190 YLTPQGGAAPOPHPPAPAFSPFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYGLDVPV 1249
 DB 1196 YLTPQGGAAPOPHPPAPAFSPFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYGLDVPV 1255

RESULT 14

AAU77114
 ID AAU77114 standard; Protein; 1255 AA.

XX AC AAU77114;

DT 05-JUN-2002 (first entry)

XX DE Human Her-2/neu polypeptide.

XX KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;

XX KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;

XX KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;

XX KW Hodgkin's lymphoma; T cell therapy.

XX OS Homo sapiens.

XX PN WO200213847-A2.

XX PD 21-FEB-2002.

XX PF 13-AUG-2001; 2001WO-US25408.

XX PR 14-AUG-2000; 2000US-0638280.

XX PR 28-SEP-2000; 2000US-0675904.

XX XX (CORI-) CORIXA CORP.

XX PI Gaiger A, Cheever MA, Hand-zimmermann S;

XX DR WPI; 2002-280741/32.

XX DR N-PSDB; ABK10730.

XX DR Inhibiting haematological malignancy development by administering

PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide

PT encoding the polypeptide, or antigen presenting cells expressing the

PT polypeptide

XX PS Disclosure; Page 71-74; 74pp; English.

XX XX

CC The invention relates to a method for inhibiting development of
 CC haematological malignancy in a patient by administering a polypeptide
 CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
 CC encoding the polypeptide. Antigen presenting cells that express the
 CC protein can also be administered. The sequences are used for inhibiting
 CC development of haematological malignancy such as acute myelogenous
 CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
 CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
 CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
 XX

SQ Sequence 1255 AA;

Query Match 97.4%; Score 6619; DB 23; Length 1255;

Best Local Similarity 96.9%; Pred. No. 0;

Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;

QY 1 MELAALCRKWLGLLALLPPGAASVCTCTDKMLRLPASFPETHLDMRLHYQCCVQVGNL 60
 DB 1 MELAALCRKWLGLLALLPPGAASVCTCTDKMLRLPASFPETHLDMRLHYQCCVQVGNL 60
 QY 61 ELTYLPTNASLSFLQDIOEVQYVLIAHNQVQVPLQRLRIVRGTLQDPEDNYALAVLNG 120
 DB 61 ELTYLPTNASLSFLQDIOEVQYVLIAHNQVQVPLQRLRIVRGTLQDPEDNYALAVLNG 120
 QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
 DB 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
 QY 181 LTLIDTNRSRACHPCSPCKGSRGWSSBESQSLTRTVACAGGCARCKGPLPTDCHEQC 240
 DB 181 LTLIDTNRSRACHPCSPCKGSRGWSSBESQSLTRTVACAGGCARCKGPLPTDCHEQC 240
 QY 241 AAGCTGPKHSDCLACLNHNSGICELHCPALVTYNTDTFESMPNPEGYTTFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACLNHNSGICELHCPALVTYNTDTFESMPNPEGYTTFGASCVTACP 300
 QY 301 YNYLSTDVSGCTLVCPLNHNSGICELHCPALVTYNTDTFESMPNPEGYTTFGASCVTACP 360
 DB 301 YNYLSTDVSGCTLVCPLNHNSGICELHCPALVTYNTDTFESMPNPEGYTTFGASCVTACP 360
 QY 361 ELEFAGCKKIFGSLAFIPESPDGPASNTAPLQEQLOVFTLEBITGYLIYSANPDSL 420
 DB 361 IQEFAGCKKIFGSLAFIPESPDGPASNTAPLQEQLOVFTLEBITGYLIYSANPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGSRISRELGSGLALIHNNHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGSRISRELGSGLALIHNNHLCFVHTV 480
 QY 481 PWDQLFRNPHQALLHTANRPEDECVGBGLACHQLCARGHCWGPQTQCVNCSQFLRGQEC 540
 DB 481 PWDQLFRNPHQALLHTANRPEDECVGBGLACHQLCARGHCWGPQTQCVNCSQFLRGQEC 540
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEVCVARC 600
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEVCVARC 600
 QY 601 PSGVKPDLSTYMPYKFWPDEGACQPCINCTHSCVDLDDKCCPAEORASPLTSIQIYKAN 660
 DB 601 PSGVKPDLSTYMPYKFWPDEGACQPCINCTHSCVDLDDKCCPAEORASPLTSIQIYKAN 660
 QY 661 SKFTGI-----TELKRRQOKIRKYTMRLLOBELTELVEPLTPSGAMPNOAMRIL 709
 DB 661 SKFTGI-----TELKRRQOKIRKYTMRLLOBELTELVEPLTPSGAMPNOAMRIL 709
 QY 710 KETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMA 769
 DB 710 KETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMA 769
 QY 770 GVGSPPVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGLSGQDLLNWCQIAKMSY 829
 DB 776 GVGSPPVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGLSGQDLLNWCQIAKMSY 835
 QY 830 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPIKWMALLESIL 889

|||||
836 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPKWMMALESIL 895
890 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIM 949
896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIM 955
950 VKCWMIDSECRPRFRELVSFSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1009
956 VKCWMIDSECRPRFRELVSFSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1015
1010 DLVDAEYLVPOQGFCCPDPAFCAGGMVHRRHSSTRSGGDLTLGLSPSEEEAPRSP 1069
1016 DLVDAEYLVPOQGFCCPDPAFCAGGMVHRRHSSTRSGGDLTLGLSPSEEEAPRSP 1075
1070 APSEGAGSDVFDGLGMAKGLQSLPHDBPLQRYSEDPTVPLPSETDGYVAPLTCSP 1129
1076 APSEGAGSDVFDGLGMAKGLQSLPHDBPLQRYSEDPTVPLPSETDGYVAPLTCSP 1135
1130 QPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVVKOVFAFGAVENPE 1189
1136 QPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVVKOVFAFGAVENPE 1195
1190 YLTPOGGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1249
1196 YLTPOGGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 15

AAR39568
ID AAR39568 standard; Protein; 1433 AA.

AC AAR39568;

DT 07-FEB-1994 (first entry)

DE Sequence of c-erbB-2 tumour antigen.

KW Tumour antigen; c-erbB-2; glycoprotein.

OS Homo sapiens.

PN W09316185-A.

PD 19-AUG-1993.

PF 05-FEB-1993; 93WO-US01055.

PR 06-FEB-1992; 92US-0831967.

PA (CETU) CETUS ONCOLOGY CORP.
(CREA-) CREATIVE BIOMOLECULES INC.

PI Houston LL, Huston JS, Oppermann H, Ring DB;

XX WPI; 1993-272889/34.

DR N-PSDB; AAQ46083.

PT New single chain Fv polypeptide binding to C-erbB-2 tumour

PT antigen - for imaging or treating breast or ovarian cancer etc.

PS Disclosure; pages 48-54; 87pp; English.

XX c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents
CC the location of a stop codon in AAQ46083.

XX Sequence 1433 AA;

Query Match 96.8%; Score 6576; DB 14; Length 1433;

Best Local Similarity 96.3%; Pred. No. 0;

Matches 1214; Conservative 8; Mismatches 22; Indels 16; Gaps 2;
QY 1 MELAALCRWGLLALLPBCAASVCTCTDMKRLPASPETHLDMLRHLYQGVQGNL 60
DB 1 MELAALCRWGLLALLPBCAASVCTCTDMKRLPASPETHLDMLRHLYQGVQGNL 60
QY 61 ELTYLPTNASLSFLODIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLODIQEVQGYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQORNPOLCYQDITLWKDIPHKQNOLA 180
DB 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQORNPOLCYQDITLWKDIPHKQNOLA 180
QY 181 LTLIDTNRSRACHPCSPCKGSRGWSESDCSLTRTVCCAGCARCKGKPLPTDCCHEQC 240
DB 181 LTLIDTNRSRACHPCSPCKGSRGWSESDCSLTRTVCCAGCARCKGKPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTORCEKSPCARVCYGLQMOYIKANSKFIGIT 360
DB 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTORCEKSPCARVCYGLQMOYIKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAFLPESPDGPASNTAPLQPEOLQVFETLEETGYLYISAMPDLSL 420
DB 361 IOEPAGCKRIFGSLAFLPESPDGPASNTAPLQPELHVFELEQITGTYLYISAMPDLSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRLSRLRELGSGLALIHNNTHLCSFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRLSRLRELGSGLALIHNNTHLCSFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQCHGCHWGPPTQCVNCSOFLRGQEC 540
DB 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQCHGCHWGPPTQCVNCSOFLRGQEC 540
QY 541 VEECEVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHKDPFCVCARC 600
DB 541 VEECEVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHKDPFCVCARC 600
QY 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIQYIKAN 660
DB 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSI ----I 665
QY 661 SKFIGI-----TELKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRIL 709
DB 661 SKFIGI-----TELKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRIL 715
QY 710 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMA 769
DB 716 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMA 775
QY 770 GVGSPVSRLLGICLTSTVQLVTLMPYGLLDHVRNRRGRGLSQDLNMCQIAKGMYSY 829
DB 776 GVGSPVSRLLGICLTSTVQLVTLMPYGLLDHVRNRRGRGLSQDLNMCQIAKGMYSY 835
QY 830 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPKWMMALESIL 889
DB 836 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPKWMMALESIL 895
QY 890 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIM 949
DB 896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIM 955
QY 950 VKCWMIDSECRPRFRELVSFSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1009
DB 956 VKCWMIDSECRPRFRELVSFSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1015
QY 1010 DLVDAEYLVPOQGFCCPDPAFCAGGMVHRRHSSTRSGGDLTLGLSPSEEEAPRSP 1069
DB 1016 DLVDAEYLVPOQGFCCPDPAFCAGGMVHRRHSSTRSGGDLTLGLSPSEEEAPRSP 1075

Qy	1070	APSEGAGSDVFDGLGMGAAGLSLPTHDPSLQRYSEDPTVPLPSETDGYVAPLTCSP	1129
Db	1076	APSEGAGSDVFDGLGMGAAGLSLPTHDPSLQRYSEDPTVPLPSETDGYVAPLTCSP	1135
Qy	1130	OPEYVNPQDVRRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPE	1189
Db	1136	OPEYVNPQDVRRPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVVKDVFAFGGAVENPE	1195
Qy	1190	YLTPOGGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV	1249
Db	1196	YLTPOGGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 09:16:23
Job time : 42.7573 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.6008 Seconds
(without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-632-652-12
Perfect score: 6796
Sequence: 1 MELALCRWGLLLALLPPCA.....TFKGTPTAENPEVLGLDVPV 1249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.virus:*
- 13: sp.vertibrate:*
- 14: sp.unclassified:*
- 15: sp.virus:*
- 16: sp.bacteriap:*
- 17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6124	90.1	1259	6	O18735
2	3109	45.7	1209	11	Q9X70
3	3080	45.3	1210	11	Q9EP98
4	2693	39.6	1165	13	Q9VH40
5	2650.5	39.0	1137	13	Q9W6F6
6	2259	33.2	1328	13	P79754
7	1993.5	29.3	1433	5	Q9BIH9
8	1871	27.5	419	4	Q9UK79
9	1739	25.6	367	11	Q8R2X1
10	1697.5	25.0	412	4	Q8WYV0
11	1677	24.7	729	15	Q86712
12	1675	24.6	567	15	Q86714
13	1610.5	23.7	962	15	Q64895
14	1602	23.6	545	15	Q85468
15	1506.5	22.2	655	11	Q9WVF5
16	1490.5	21.9	643	11	Q9ERV6

17 1250 18.4 1193 5 Q9Y1X8
18 1176.5 17.3 1368 5 Q23821
19 1148 16.9 1717 5 Q26566
20 1126 16.6 527 13 Q90836
21 1001.5 14.7 478 11 Q9ESE0
22 942.5 13.9 599 13 Q9FSH2
23 906 13.3 165 4 Q14256
24 887 13.1 176 11 Q923V5
25 806.5 11.9 346 13 P11776
26 778 11.4 435 5 Q8SZW1
27 754.5 11.1 311 13 Q99162
28 744.5 11.0 1362 13 Q9PVZ4
29 734 10.8 331 4 Q9BUD7
30 723 10.6 149 6 Q9BG66
31 713 10.5 1671 5 Q9NJV5
32 686 10.1 1418 13 Q93457
33 682.5 10.0 1368 13 Q8UM85
34 661.5 9.7 1369 13 Q8UM86
35 656 9.7 1358 13 Q73798
36 648.5 9.5 1472 5 Q9USA8
37 643 9.5 1412 13 Q8UM84
38 633.5 9.3 1418 13 Q8UM83
39 619 9.1 1245 13 Q9YGH8
40 617.5 9.1 2144 5 Q9VD94
41 617 9.1 1371 11 Q9QVW4
42 586.5 8.6 1055 11 Q54967
43 586 8.6 1036 4 Q07912
44 574.5 8.5 1091 4 Q9UMQ4
45 565.5 8.3 981 15 Q92809

ALIGNMENTS

RESULT 1
O18735 ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Erbb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]_TaxID=9615;
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbb-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1; -
DR HSSP; P11362; 1FGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD0000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00018; EF HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferrase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match	90.1%;	Score 6124;	DB 6;	Length 1259;
Best Local Similarity	89.5%;	Pred. No. 0;		
Matches 1132;	Conservative	42;	Mismatches	69;
			Indels	22;
			Gaps	4;
Qy	1	MELAAACRWGGLLALLP	PGAAASTQVCTGTDMKRLPASPETHLDMLRHLVGGQVQGNL	60
Db	1	MELAAACRWGGLLALLP	PSGAAGTQVCTGTDMKRLPASPETHLDMLRHLVGGQVQGNL	60
Qy	61	ELTYLPTNASLFLQDIQ	EOGVYVLIAHNOVROVPLORLRIVRGTQLFEDNYALAVLDNG	120
Db	61	ELTYLPTNASLFLQDIQ	EOGVYVLIAHNOVROVPLORLRIVRGTQLFEDNYALAVLDNG	120
Qy	121	DPLNNTTPTVTCASPG	GLRELQRLSLTEILKGGVLIQNPOLCYQDTILMKDIFHKNNOLA	180
Db	121	DPLEGGIPAPGAAGGL	RELQRLSLTEILKGGVLIQNPOLCHQDTILMKDVFHKNNOLA	180
Qy	181	LTLIDTNRSRACHPC	SMKGRSCWGESSEDCQSLTRTVAGGCARCKGBLPTDCCHEQC	240
Db	181	LTLIDTNRSRACHPC	SMKGRSCWGESSEDCQSLTRTVAGGCARCKGBLPTDCCHEQC	240
Qy	241	AAGCTGPKHSDCLAC	LHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Db	241	AAGCTGPKHSDCLAC	LHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTSCP	300
Qy	301	YNYLSTDVGSCTLV	CPHLHNOEVTAEQDQRCCKSKPCARVCYGLGMQYIKANSKFITGIT	360
Db	301	YNYLSTDVGSCTLV	CPHLHNOEVTAEQDQRCCKSKPCARVCYGLGMHLEHREAVRVTSAN	360
Qy	361	ELEFAGCKKIFGSLA	PIFESFDGDPASNTAPLQEQLOVFTLBEITGYLISAWPDSLP	420
Db	361	IQEFAGCKKIFGSLA	PIFESFDGDPASNTAPLQEQLRVFEALBEITGYLISAWPDSLP	420
Qy	421	DLVSFONLQVIRGR	LHNGAYSLTLOGLIGLSWGLRSLRGLSLALIHNTHLFCFVHTV	480
Db	421	NLSVFQNLVIRGRV	LHDGAYSLTLOGLIGLSWGLRSLRGLSLALIHNRNLCFVHTV	480
Qy	481	PWQDLFRNPHOALL	TNRRPEDECVGEGGLACHQLCARGHGWGPGPTQCVNCSQFLRQEC	540
Db	481	PWQDLFRNPHOALL	HSANRPEEECVGEGGLACYP-CAHGHGWGPGPTQCVNCSQFLRQEC	539
Qy	541	VEECRVQLGLPREV	NARHCLPCHPECPQNGSVTCFGEADQCVACAHKDPFPCVARC	600
Db	540	VEECRVQLGLPREV	NKDYRLCPCHSECPQNGSVTCFGEADQCVACAHKDPFPCVARC	599
Qy	601	PSGVKPDLSYMPI	WKPFDEEGACOPCPINCTHSCVDLDKGPACQASPLTSTQYIKAN	660
Db	600	PSGVKPDLSYMPI	WKFADEEGTCOPCPINCTHSCADLDKGPACQASPLTSTQYIKAN	654
Qy	661	SKFIGI-----	TELKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRIL	709
Db	655	AAVVGILLAVVVL	GLILKRRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRIL	714
Qy	710	KETELRKVKVLG	SAGFTVYKGIWIPGENVKIPVAIKVLENTSPKANKEILDEAYMA	769
Db	715	KETELRKVKVLG	SAGFTVYKGIWIPGENVKIPVAIKVLENTSPKANKEILDEAYMA	774
Qy	770	GVGSPYVSRLLIG	ICTSTVOLVTOLMPYGCILLDHVRENRRGLSGQDLLNWCQIAKMSY	829
Db	775	GVGSPYVSRLLIG	ICTSTVOLVTOLMPYGCILLDHVRENRRGLSGQDLLNWCQIAKMSY	834
Qy	830	LEDVRLVHRDLA	ARNVLKGNHVKITDIFGLARLLDIDETEHADGKGVPKKNMALESIL	889
Db	835	LEDVRLVHRDLA	ARNVLKGNHVKITDIFGLARLLDIDETEHADGKGVPKKNMALESIP	894
Qy	890	RRRETHOSDVMS	YGVTVWELMTFCAPKYPDIPAREIPOLLEKGERLPQPPICITDVTYIM	949
Db	895	PRRTHOSDVMS	YGVTVWELMTFCAPKYPDIPAREIPDLLEKGERLPQPPICITDVTYIM	954
Qy	950	VKCMIDSECRPR	PRELVSFBSRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDDMG	1009
Db	955	VKCMIDSECRPR	PRELVAEFSRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDDMG	1014
Qy	1010	DLVDAEYLVPOQGF	FCPEPTPCAGGTAAHRRHSSSTRNGGELTLGLEPSEEPKSP	1069

Db	1015	DLVDAEYLVPOQGF	FCPEPTPCAGGTAAHRRHSSSTRNGGELTLGLEPSEEPKSP	1074
Qy	1070	APSEGAGSDVFDG	DLGMAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSP	1129
Db	1075	APSEGAGSDVFDG	DLGMAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSP	1134
Qy	1130	QPYVNVQPDVVR	PPPPREGPLPAARPAGATTLER-----AKTILSPGKNGVVKDVFARFGA	1184
Db	1135	QPYVNVQPDVVR	PPPPREGPLPAARPAGATTLERPKTILSPGKNGVVKDVFARFGA	1194
Qy	1185	VENPEYLTPOGGA	APQHPHPPAFSPAFDNLVYWDQPPPERGAPSTPKGTPTAENPEYLG	1244
Db	1195	VENPEYLTPOGGA	APQHPHPPAFSPAFDNLVYWDQPPSERGSPSTTEGTPTAENPEYLG	1254
Qy	1245	LDVPPV	1249	
Db	1255	LDVPPV	1259	
RESULT 2				
Q9QX70				
ID	Q9QX70	PRELIMINARY;	PRT;	1209 AA.
AC	Q9QX70;			
DT	01-MAY-2000	(T-EMBLrel. 13, Created)		
DT	01-MAY-2000	(T-EMBLrel. 13, Last sequence update)		
DT	01-JUN-2002	(T-EMBLrel. 21, Last annotation update)		
DE	Epidermal growth factor receptor.			
GN	EGFR.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OC	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FISHER; TISSUE=LIVER;			
RX	MEDLINE=9025888; PubMed=2342466;			
RA	Petch L.A.; Harris J.; Raymond V.W.; Blasband A.J.; Lee D.C..			
RA	Earp H.S.;			
RT	"A truncated, secreted form of the epidermal growth factor receptor is			
RT	encoded by an alternatively spliced transcript in normal rat tissue."			
RL	Mol. Cell. Biol. 10:2973-2982(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FISHER; TISSUE=LIVER;			
RA	Petch L.A.;			
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FISHER; TISSUE=LIVER;			
RA	Guttridge K.; Dawson T.L.; Earp H.S.;			
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; M37394; AAF14008.1; -.			
DR	HSP; P11362; IFGK.			
DR	InterPro; IPR000494; EGFR L domain.			
DR	InterPro; IPR000719; Euk_pkinase.			
DR	InterPro; IPR002174; Furin-like.			
DR	InterPro; IPR001245; Tyr_pkinase.			
DR	Pfam; PF00757; Furin-like; 1.			
DR	Pfam; PF00069; pkinase; 1.			
DR	Pfam; PF01030; Recep_L domain; 2.			
DR	PRINTS; PR00109; TYRKINASE.			
DR	ProDom; PD000001; Euk_pkinase; 1.			
DR	SMART; SM00261; FU; 3.			
DR	SMART; SM00219; TYRK; 1.			
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.			
DR	PROSITE; PS00011; PROTEIN KINASE DOM; 1.			
DR	PROSITE; PS00109; PROTEIN KINASE TYR; 1.			
KW	ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.			
SQ	SEQUENCE 1209 AA; 134891 MW; 96FEET66C1B7773 CRC64;			
Query Match		45.7%;		Score 3109; DB 11; Length 1209;
Best Local Similarity		50.0%;		Pred. No. 8.7e-226;

Best Local Similarity 46.8%; Pred. No. 3.5e-191; Matches 533; Conservative 165; Mismatches 358; Indels 83; Gaps 26;	
QY	161 LCYQDTILWKDIFHKQNKALTLIDNRSACHPCSPCKGSRGWESSEDQSLTRTVIC 220
DB	3 LCFAADIHWQDIVRNWPASNFLLVPTNGSGGCRCHKSCGT-RCWGPTEHNCQTLTKTVIC 61
QY	221 AGGC-ARCKGPLPTDCHCQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDTF 279
DB	62 AEQDGRGCVGYVSDCHREHAGCGSPKDTDFACWPNFDSGACVTCQPTQFVYNTTF 121
QY	280 ESMNPPEGYTTFGACSVTACPVNYLSTVGSCTLVCPHNNQEVTAEDGTQRCCKSKPCA 339
DB	122 OLEHNHNAKYTYCAFCVKXCPHNFV-VDSGSCVRACPSKMEV-EENGIMCKPCPTDICEP 179
QY	340 RUCYGLGMQYIKANSKFIIGITELE-FAGCKKIIGSLAFIPESDFGDPASNTAPLQEQLO 398
DB	180 KACDGGTGS-LVSAQTVDSSNIDKFNCTKNGNLIFLVTHGDPYHTIAINPEKLN 238
QY	399 VFETLEETGYLYISAWPDSLPDSVFQNLQVIRGILHNGAYSLLTLQGLIGISWLGRLSL 458
DB	239 IFQTVREITGYLNIQSPENMTDFRVFSLVTIGGRALYSGLSLLILKQGGITSLQFQSL 298
QY	459 RELGSLAIHHNTHLCFVHTVPMWQDLFRNPHQALLHTANRPEDECVGEGLAGHQICARG 518
DB	299 KOISAGNIYITDNSLNCYVHTVNTSLFSTPSQKTVIHRNKAENCTADGMVCNELCSSD 358
QY	519 HCWGPPTQVNCVSOFLRGQECVEECRVLOGLPREYVNAHCLPCHPEQCP-QNGSVTGF 577
DB	359 GCMGPGDQCLSKRIFRGKTCIESNLYDGEFREFANGSVCMCEQDQCEKMDNMITCY 418
QY	578 GPREADCVACAHKDPFPFCVARGPSGVKPDLSYMPIWKFDPDEGACQPCPINCHSCVDL 637
DB	419 GPGPDHCTCFHFKDGNVCVEKCPDGLQANSF--IFKYADEBRECHPCHPNCTQCRGP 476
QY	638 DDKGC-----PAEQASPLTSIQYIKANSKFI-----GIT--ELKRRQOKIRKYT 690
DB	477 ASHDCIYYPWTRQSTLPQHAR-TPLIAAGVI--GGLFIIIMGLTFAVYVRRKSIKKRA 533
QY	681 MRELLOETELVEPLTSGAMPNQAOMRILKETELRKVKVLGSGAFCTVYKGIWIPGENV 740
DB	534 LRRL-ETELVEPLTSGTAPNQAORILKETELRKVKVLGSGAFCTVYKGIWIPGETV 592
QY	741 KIPVAIKVRENTSPKANKEILDEAYVMAGVGPVYVSRLLIGICTSTVQLVLTQLMYPVGL 800
DB	593 KIPVAIKILNETGPRANVEFMDEALIMASMDHPLVRLGVCLSPITQLVLTQLMHPGCL 652
QY	801 LDHVRNRRGLSQDILLNWCQIAKMSYLEDVRLVHRDLAARNVLVKSNNHVKITDFGL 860
DB	653 LDYVHEHKDNGISQILLNWCQIAKGMVYLEERLVRDLAARNVLVKSNNHVKITDFGL 712
QY	861 ARLLDIDETEHADGGKVPILKMMALESILRRRRETHQSDVMSYGVYVWELMTFCAPYDGI 920
DB	713 ARLLGEKEYNADGGKVPILKMMALECIIHYRKETHQSDVMSYGVYVWELMTFCGKPYDGI 772
QY	921 PAREIPDLLEKGERLPQPPCTIDVYVMVVKMWIDSECRPRFELVSEFSRMRARDPQRF 980
DB	773 PTEIPDLLEKGERLPQPPCTIDVYVMVVKMWIDADSREKFKELAAEFMRARDPQRY 832
QY	981 VVJQNEP-LGPASPLDSTFVRSLLDDDDGDLVDABEYLVPOQFPDPPAPAGAGMWHH 1039
DB	833 LVIQGDDRMKLPSPNDSKFFQNLLEEDLEDMDMAEYLVLP-QAFNIPPIYTSRTIDS 891
QY	1040 RHRSSTRSGGDLTLGLEPSEEAAPRS--PLAP-SEGAGSDVDFDGLGMAAGKQSLP 1096
DB	892 NRNQFYVRDGYAAEQGV-PMPYRACGCIIPAPVPAQGATAEIPEDTCCNGTKURKQVATL 950
QY	1097 THDPSPLQRYSEDPTVPLPS-----ETDGYVAPLTCSPQPEYVNPQDVRPQPPSPREG 1149
DB	951 AKEDSSTQRYSDPTVPIPERVIRGELDEGDYNTPMRDKPKTDYLNPEVENPFVSRKNG 1010
QY	1150 PLPAA-RPAGATLERAKTSLSPGNKVGVKDF-----AFGAVENPEYLTTPQGAAPQ 1200

DB	1011 DLQAVDNPEYHN-----APNQCPKADEYVNEPLYLNTFANTLENAEYL-----K 1055
QY	1201 PHPPAPAFSAPFONLYWQDQDPERGA--PSPFKGTPT-----AENPEYL 1243
DB	1056 NNLPEKAKAFONPDYWNHSLPPRSTLQHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1114
RESULT 6	
ID	P79754 PRELIMINARY; PRT: 1328 AA.
AC	P79754;
DT	01-MAY-1997 (Tremblrel. 03, Created)
DT	01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE	Erb83;
OS	Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC	Tetraodontidae; Takifugu.
OX	NCBI_TaxID=31033;
RN	[1]_TaxID=31033;
RP	SEQUENCE FROM N.A.
RX	MEDLINE=99177347; PubMed=10077531;
RA	Gellner K., Brenner S.;
RT	"Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT	rubripes.";
RL	Genome Res. 9:251-258 (1999).
DR	EMBL: AF056116; AAC34391.1; -.
DR	HSSP: P11362; IFGK.
DR	InterPro: IPR000494; EGFR L domain.
DR	InterPro: IPR000719; Euk_pkinase.
DR	InterPro: IPR002174; Furin-like.
DR	InterPro: IPR001245; Tyr_pkinase.
DR	Pfam: PF00757; Furin-like; 1.
DR	Pfam: PF00069; pkinase; 1.
DR	Pfam: PF01030; Recep_L domain; 2.
DR	ProDom: PD000001; Euk_pkinase; 1.
DR	SMART: SM00261; FU_3.
DR	SMART: SM00219; TyrKC; 1.
DR	PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW	ATP-binding; Transferase.
SQ	SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;
Query Match 33.2%; Score 2259; DB 13; Length 1328; Best Local Similarity 40.1%; Pred. No. 1.7e-161; Matches 516; Conservative 157; Mismatches 417; Indels 196; Gaps 35;	
QY	9 WGLLLALLPP--GAASTQ---VCTGTDMLRLPASPEHLDMLRHLQYQGVVQGNLEL 62
DB	4 WRLILNCVASRLRAASSQTQEA VCPGTQNGLSSTGSEQNYNLNKORYKCEIIMGNLEI 63
QY	63 TYLPTNWSLFLQDIQEVGYVLIHNRVQVPLQRLRIVRGTLQFEDNYVALVLDNGDP 122
DB	64 TQIESNWFDFLKTIREVTGYVLIAMHFQEIPLGQLRVIRGNSLYRERREALSVFLN--- 120
QY	123 LNNTPVTGASPGLELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHKQNKALTL 182
DB	121 ----YPKDG--PSGLNQLGLMNLTEILDGQVQIINNKLRYGVPWVYWRDII--RNNDAPIE 173
QY	183 LIDNTRSRACHPCSPCKGSRGWESSEDQSLTRTVCAAGC-ARCKGPLPTDCCHEQCA 241
DB	174 IQFNGERGVCVH---KSC-GNYCWGPGKQDQQLTKTVCAPOCNDRCFGTSPDCCHECA 229
QY	242 AGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDTTFESMPNPEGYTTFGACSVTACPY 301
DB	230 AGCKGLDTCFACRLFNDSGACVPCQPTLIYKQTFQMETPNPAKYQYGSICVSCQPT 289
QY	302 NYLSTDVGSCTLVCPHNNQEVTAEDGTQ-CEKCSRPCARVCYGLHMYKANSKFIKIT 360
DB	290 HFV-VDGSSCVSPPPDKMEV--ERGSORQCELCGLCPKVCSEGTGAE-----QRQTVDS 342

```
QY 361 ELE-FAGCKKIFGSLAFLESPDGPASNTAPLOPELOQVFEETLEBITGYLISAWPDSL 419
Db 343 NIDSFNCTKIOGSLHFLVTGILGDDFKXVPPDLAKKLEVFVRVREITDILNQSFPKEL 402
QY 420 PDLVSFQNLQVIRGRILHNGAYSLTLQGLGSLWGLRSLRSLGSLALIHNNHLCFVHT 479
Db 403 NDLVSFSLTTIOGRSLFKRSLVMWRIPLTLSGLRSLREISDGSVYSQNAHLCHYHT 462
QY 480 VPWDQLFRNPH-QALLHTANREDECVGGLACHQLCARGHCWGPQPTQCVNCSQFLRQ 538
Db 463 VNMTOFLFRGSRVANSLSNRPMACVADGRVCDPLCSDSGCWGPGPDQCLSCRNSRHG 522
QY 539 ECVEECRLVQLPREVYNARH-CLPCHPECPONGSVTCFGEADOCVACAHYKDPFCV 597
Db 523 TVAGCHFNHGIIPREFAGLNGVCVACHPECKPOTGASCTGPGADECMACTFRDGPYCM 582
QY 598 ARCPGKVPDLSPYMIWKPDEEGACQPCPINCTHSCVDLDKKGCPAEQR----ASPLTS 653
Db 583 SSCPAGVN-DGEKGLIFKFPNREHCEPCHONCTQCGSGPLNCLCEAARLTISSQITG 641
QY 654 IOY-IKAN-----SKFIGITELKROQKI-RKYMTRILLQETELVEPLTPSGAMPNQAO 705
Db 642 IALGVPAGLIFCLVLFLLGM--LYHRGLAIRKRAMRYLESSESPEPLGP-GEKTKVH 698
QY 706 MRILKTELKVKVLCGSAFGTVYKIGWIPDGENVKIPVAIKVIRENTSPKANKETLDEA 765
Db 699 ARILKPSDLRKIKPLGSGVGTGKFWIPGETVKIPVAIKTIQSSSGRQTFETIDHL 758
QY 766 YMVAGVSPYRSLGILCITSTVQLVTQMPYGCLLDHVRENRRGLSGDOLLNWCQIAK 825
Db 759 LSMGSLDHPYIVRLGICPGTCLQLVTQLSSHGSLLEHROKHTSLDPQRLNWCQIAK 818
QY 826 GHSYLEDVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMAL 885
Db 819 GMYLLEHVRVHKNLAARILLKNDYQVQISYGVADLLYPDKKYVSETKTPKMMAL 878
QY 886 ESILRRRFTHQSDVMSYGVTVWELMTFFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDV 945
Db 879 ESILFRYTHQSDVMSYGVTVWEMSGFAEPYASVQBPQVPSVLEKGERLSQPAICTIDV 938
QY 946 YMIMVKWMIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLED 1005
Db 939 YMVVKWMIDENIRPTFKELASDFTRMARDPRLVIRMEG-----ED 982
QY 1006 DMGDLVDAEEVLVPOQGFCDPAPGAGGMVHHRSSRSTSGGGDLTLGLEPSEEEAP 1065
Db 983 SGMGEFL-----RRGSR--GLLEADLEDEEE-- 1008
QY 1066 RSLAPSEGAGSDVFDGDLGMG--AAKGLQSLPTHDRSPLO-----RYSEDPTV 1112
Db 1009 -----GLGDRFATPSLQSPSWSTSPQINSYMWVMTQURYD----- 1044
QY 1113 PLPSETDGYVAPLTCSPQ-EYVNO-----PDVRPQPPSPREGPL--PAAR 1155
Db 1045 -FAVSGGHIGYLPSPSPVDITRQLWYORSRLSSVRTLPRDSAFRRSREAELEDGQAO 1103
QY 1156 PAGATLERAKTLSPCKNGVWVDVAFGGAVENTPEYLTPOGGAPOPHPPAPSPADNLY 1215
Db 1104 CAGIFRVR-----FGSERGN-----PQGG----- 1122
QY 1216 YWDQPPPERGAPPSTFKGTPTAENPE 1241
Db 1123 --QQRKLSTASSPSSFKTWADEDEE 1146

RESULT 7
Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
```

```
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]_TaxID=7165;
RC SEQUENCE FROM N.A.
RC STRAIN=SU4;
RA Lyceett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TK; 1.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER.
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.3%; Score 1993.5; DB 5; Length 1433;
Best Local Similarity 32.2%; Pred. No. 2.2e-141;
Matches 466; Conservative 194; Mismatches 396; Indels 391; Gaps 37;

QY 26 CTGDMKLRRLRASPETHLDMLRHLVGGCQVQVGNLELTYLPTNASLSFLQDIOEVGYVL 85
Db 1 CIGTNGRMSVPANREYHYKNLDRYTNCTYVDNLEITWIAITDLNLFQIHREVGYVL 60
QY 86 TAHNQVROVPLQRLRIVRGTLF-----EDNALAVLDNGDPLNNTPTVTGASPGGLREL 140
Db 61 ISLYDLQVILPRLQIIGRTTFKLNKWEAYGLFV-----SFSHMTL 104
QY 141 QLRSLTEILKGGVLIQRNPQLCYQDTILWKDI-FHKNNQALTLIDNRSRACHPCSPMC 199
Db 105 ELPALRDLILGSGVGFNNVNLCHKMSINWEEILLAPQTSQMYTFNFPSSPERVCPCHPSC 164
QY 200 KGSRCWGESSEDCQSLRTRVCAGGCA--RCKGPLPTDCCHQCAAGCTGPKHSDCLACUH 257
Db 165 EVG-CWGEAHNCORFSLKNCSPQSGRCFCGPKPRECCHLFCAGCTGPTQSDCLACKN 223
QY 258 FNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNYLSLTVGSCSLVCP 317
Db 224 FYDDGVCKQCEPQMIQIYNPTNYFWPNPDGKTAYGATCVRKCP-BHLLKDNAGACVKKCPK 282
QY 318 HNQEVTAEQGTQRCCKSKPCARVCYGLGMQYIKANSKFIGITELEFAGCKKIFGSLAF 377
Db 283 GMPQNSE-----CVPCKGVCPKTCPEGI----VHSDNIG----NYKDCITIEGSL 329
QY 378 PESDGDPAISNT-----APLQPEQLQVETLEEITGYLYISAWPDSLPLSQFQNIQ 429
Db 330 DQSPDGFQOQVYTNFSGPRYIKIDPRLLEVSTVKEITGFINQIAHPNFTLTINFRNLE 389
QY 430 VIRGRILHNGAY-SLTLOGLGISMLGRSLRSLGSLALIHNNHLCFVHTVPMQOLFRN 488
Db 390 VVGRQLKENLFASVYIVKTSLSLEKSLKRVNMSGVILENSDLCFVEDIDWSEIKKS 449
```



```
QY 489 PHOALLHTANRPEDECVGEGLACHQLCARGCHGPGTQCVCNCSQFLRGQECVCECVRLQ 548
Db 450 SDHEVMVQKRNATECHEEGMECSQCSKAGCWKGPEQCLCKNVKYKCKLDSCK--- 506
QY 549 GLPREY-VNARHCLPCHPEQCPONGSVTCFGEADQCVACAHYKDPFCVAPCP----- 601
Db 507 SLPLYSVDSKTCODCHQCKD-----FCYGPNEGNCSCMNKVDGRFCVACPTTKHAM 561
QY 602 -----SGVKPDLSPYPIWKPPD----- 618
Db 562 NGTCINCHKTCVGRGRDTHIADPGCISCDKAIIGSDAKIERCLMKDESPDGYSDYVL 621
QY 619 -EEG----- 621
Db 622 QEEGLKQLSGKAVCRKCHPRCKKCTGYPHEQFCQECTGYKKGEOCEDECPQDFYANEE 681
QY 622 --ACQCPINCT-----HSCVDL-----KGCPAQ----- 646
Db 682 TRICLPCHQECROCHGLGDHHCRLNKLFEQDPYDNATTTVCVSNCPASHPYKRPQEA 741
QY 647 -----RASPLTSIOYIKANSKF-----IGITELKRROQKIRKYTM 681
Db 742 KGIGPYCSADSMOSGLRIEFTQVKIVGSMVALILLCVVGTAFLVFSRHKKKDAVKM 801
QY 682 RRLLOETELVEPLTPSGAMPNQMRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVK 741
Db 802 TMLAGCEDSEPLRPSNVGNLTKLRIKEAIRRGVLGNGAFGRVFKGVWMPGEGSVK 861
QY 742 IPVAIKVLRNTPSKANKEILDEAYNAGVGSFVSRLLGICLTSTVOLTQMLPYCCLL 801
Db 862 IPVAIKVLMEMSGESSEKFELEAYIMASVEHPNLLKLLAVCMTSQMLLTQMLPLCCLL 921
QY 802 DHVRENRLGSLGODLNNWCQIAGMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLA 861
Db 922 DYVRNKKDKIGSKALLNWSQIARGMAYLEERRLVHRDLAARNVLTQPCVKLTVEGLA 981
QY 862 RLIDIDETEHADGGKVPKIMMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIP 921
Db 982 KLLDFDSDEYRAAGGKMPKWLALCEIRHVRFTSKSDVWAFGTIWEITYGARPYENVP 1041
QY 922 AREIPDLLEGERLPOPICTIDVYIMLVKCMWIDSECRPRFRELVSERFMRARDPORFV 981
Db 1042 AKDVPELIEIGHKLPPQDICSLDVYICLLSCWLDADARPTFKQLAETFAEKARDPGYL 1101
QY 982 VIONEDLGPASPLDSTFYRSLLEDDDDMDLV----- 1012
Db 1102 MI-----PGDKFMRLPSTYNQDEKDLIRTLAPVMAAAAAAAGASNVDPSTIA 1152
QY 1013 DABEYLVPQOGFFCPDPAPAGGMVHHRSSSTRSGGDLTLGLEPSEBEAPRS----- 1067
Db 1153 ETDEYLQPKTRPSIMLPGSA-----VEPS-DEMPKSLRYCK 1188
QY 1068 -PLAP---SEGAGSDVFDGDLGMAKGLQSLPHTDPSLQRYSEDPTVPLPSETDGYVA 1123
Db 1189 DPLKPDDETDGHEV-----GVGGIR-----LNLPLDEDDYLM 1222
QY 1124 PLTCSPOPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVWVDVAFGG 1183
Db 1223 P-TCQSQ-----NQS-----TPG-----YMDLIGVPA 1243
QY 1184 AVNPEYL-----TPQGGAAPOHPPPAPSPAFDNLYWDDQDPPRGAPSTPKGT 1234
Db 1244 SVDNPEYLMGSTQAIAGLAGSMG--PHTTP-----PPNTNGM 1280
QY 1235 PTAENPE 1241
Db 1281 PTHQHSQ 1287
```

RESULT 8

Q9UK79

ID Q9UK79 PRELIMINARY; PRT; 419 AA.

AC Q9UK79;

```
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 27.5%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 6.2e-133;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MELAAACRGLLALLPPGAASSTOVCCTGDMKLRPASPTHLDMLRHLVGGCOVVOGNL 60
Db 1 MELAAACRGLLALLPPGAASSTOVCCTGDMKLRPASPTHLDMLRHLVGGCOVVOGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVLYIAHNOVQVPLQRLIVRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVLYIAHNOVQVPLQRLIVRGTLQFEDNYALAVLDNG 120
QY 121 DPLNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFHKNOLA 180
Db 121 DPLNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFHKNOLA 180
QY 181 LTIDNRSRACHPCSPMCKSGSCWGESSEDCOSLRTVCAGGCARCKPLPTDCCHEQC 240
Db 181 LTIDNRSRACHPCSPMCKSGSCWGESSEDCOSLRTVCAGGCARCKPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNHOEVTAEDGTQRCCKSPCARVCYGL 345
Db 301 YNYLSTDVGSCTLVCPLNHOEVTAEDGTQRCCKSPCARVCYGL 345

RESULT 9
Q9R2X1 PRELIMINARY; PRT; 367 AA.
AC Q9R2X1;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
```

```
DR EMBL; BC027080; AAH27080.1; -.
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match      25.6%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 4.8e-133;
Matches 323; Conservative 11; Mismatches 23; Indels 0; Gaps 0;

QY 883 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 942
DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 60

QY 943 IDVYIMVWKWIMIDSECRPRFRELSEFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1002
DB 61 IDVYIMVWKWIMIDSECRPRFRELSEFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 120

QY 1003 LEDDDMGDLVDAEYLVPOQGFCDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEE 1062
DB 121 LEDDDMGDLVDAEYLVPOQGFCDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEE 180

QY 1063 EAPRSPPLAPSEGAGSDVFDGLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYV 1122
DB 181 EAPRSPPLAPSEGAGSDVFDGLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYV 240

QY 1123 APLTCSPOEYVNPQDVRPQPSREGPLPAARPAGATLERAKTLPSPGKNGVVKDVFAFG 1182
DB 241 APLTCSPOEYVNPQDVRPQPSREGPLPAARPAGATLERAKTLPSPGKNGVVKDVFAFG 300

QY 1183 GAVENPEYLTPOGGAAPQHPAPAFDNLVYNDODPPERCAPPSTFKGTPTAENPEY 1242
DB 301 GAVENPEYLTPOGGAAPQHPAPAFDNLVYNDODPPERCAPPSTFKGTPTAENPEY 360

QY 1243 LGLDVVPV 1249
DB 361 LGLDVVPV 367

RESULT 10
Q8WYV0 PRELIMINARY; PRT; 412 AA.
AC Q8WYV0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Hypothetical 44.7 kDa protein.
GN PP3659.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF318349; AAL55856.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3E27D2BC CRC64;

Query Match      25.0%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 7.8e-120;

QY 569 PQNGSVTCFGEADQCAVCAHYKDPFVCVARCPGVKPDLSYMPIWKFLPESGACQPCPI 628
DB 1 PQNGSVTCFGEADQCAVCAHYKDPFVCVARCPGVKPDLSYMPIWKFLPESGACQPCPI 628
```

```
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 883 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 942
DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 60

QY 943 IDVYIMVWKWIMIDSECRPRFRELSEFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1002
DB 61 IDVYIMVWKWIMIDSECRPRFRELSEFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 120

QY 1003 LEDDDMGDLVDAEYLVPOQGFCDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEE 1062
DB 121 LEDDDMGDLVDAEYLVPOQGFCDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEE 180

QY 1063 EAPRSPPLAPSEGAGSDVFDGLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYV 1122
DB 181 EAPRSPPLAPSEGAGSDVFDGLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYV 240

QY 1123 APLTCSPOEYVNPQDVRPQPSREGPLPAARPAGATLERAKTLPSPGKNGVVKDVFAFG 1182
DB 241 APLTCSPOEYVNPQDVRPQPSREGPLPAARPAGATLERAKTLPSPGKNGVVKDVFAFG 300

QY 1183 GAVENPEYLTPOGGAAPQHPAPAFDNLVYNDODPPERCAPPSTFKGTPTAENPEY 1242
DB 301 GAVENPEYLTPOGGAAPQHPAPAFDNLVYNDODPPERCAPPSTFKGTPTAENPEY 360

QY 1215 YVWD-QDPPER-----GAPPSTFKGTPTAEN 1239
DB 361 YVWD-QDPPER-----GAPPSTFKGTPTAEN 410

RESULT 11
Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian tous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnsson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSP; P03322; 1A6S.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro_M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFE1D63 CRC64;

Query Match      24.7%; Score 1677; DB 15; Length 729;
Best Local Similarity 54.0%; Pred. No. 6.5e-118;
Matches 352; Conservative 69; Mismatches 125; Indels 106; Gaps 14;

QY 569 PQNGSVTCFGEADQCAVCAHYKDPFVCVARCPGVKPDLSYMPIWKFLPESGACQPCPI 628
DB 1 PQNGSVTCFGEADQCAVCAHYKDPFVCVARCPGVKPDLSYMPIWKFLPESGACQPCPI 628
```

```

Db 141 PEETATPKTGP--DHCMKCAHFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHP 197
QY 629 NCHTSCVDLDDKCPAEQASPLTSIQYIKANSKF-----IGITELKRROOKIRKYTWIR 682
Db 198 NTRGCKGGLGECPP---NGSKTPTSAAAGVVGGLCLLVVGGIGLYLRRHIVRKTILR 254
QY 683 RLLOETELVEPTTPSGAMPNQAMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKI 742
Db 255 RLLOERELVEPTTPSGEAPNQAHRLILKETEFKVKVLGSGAGFTVYKGLWIPGEKVKI 314
QY 743 PVAIKVLRENTSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVQLTQMLPYGCLLD 802
Db 315 PVAIKELREATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLTQMLPYGCLLD 374
QY 803 HVRENRRGLGSDLLNWCQIAKMSYLEVDRLVHRDLAARNVLKSPNHVKITDFGLAR 862
Db 375 YIREHKDNI--GSGYLLNWCQIAKMSYLEVDRLVHRDLAARNVLKSPNHVKITDFGLAR 434
QY 863 LLDIDETEHADGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGI 922
Db 435 LLGADEKEYHAEGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGSKPYDGI 494
QY 923 REIPDLLEKGERLPQPPICITIDVYIMVVKCWMIDSECRPRFRELVSFMRMARDPQRFV 982
Db 495 SEISSVLEKGERLPQPPICITIDVYIMVVKCWMIDADSRPKFRELIAEFSKMARDPPEYLV 554
QY 983 IQ-NEDLGASPLDSTFYRSLLEDDMDGLVDAEEYLVPOQGFCCPDAPGAGGMVYHRH 1041
Db 555 IODERHULPSPDTSKFTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 598
QY 1042 RSSTRSRGGDLTLGLEPSEEAAPRSP-----APSEGAGSDVDFDGLGMAAKGLQSLP 1096
Db 599 NSPST-----SRTPLLSLSATSNNATNCID-----RNGQHP 632
QY 1097 THDPSPLORYSEDPVLPSET--DGYVAPLTCSPQPEYVQDVRPQPPSPREGPLPAA 1154
Db 633 VREDSFVQRYSSDPTGNFLEESIDGFL-----PAPEYVQ--LMPKKPS----- 675
QY 1155 RPAGATLERAKTSPCKNGVVKDVF-----AFGAVENPEYL 1191
Db 676 -----TAMVQNIYNNISLTAISKLPMDSRYSQNSHSTAVDNPEYL 715

RESULT 12
Q86714
ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

```

```

DR PROSITE; PS00107; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 24.6%; Score 1675; DB 15; Length 567;
Best Local Similarity 54.6%; Pred. No. 6.3e-118;
Matches 351; Conservative 67; Mismatch 119; Indels 106; Gaps 14;

QY 578 GPEAQCVCACAHYKDPFPCVACRPSGVKPDLSYMPYIWKFPDEGACQPCPINCTHSCVDL 637
Db 1 GP--DHCMKCAHFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHPNCRTCKGP 57
QY 638 DDKGCPAEQASPLTSIQYIKANSKF-----IGITELKRROOKIRKYTWIRRLLOQETELV 691
Db 59 GLEGGP---NGSKTPTSAAAGVVGGLCLLVVGGIGLYLRRHIVRKTILRRLLQRELV 114
QY 692 EPLTPSGAMPNQAMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRE 751
Db 115 EPLTPSGEAPNQAHRLILKETEFKVKVLGSGAGFTVYKGLWIPGEKVKIPVAIKELRE 174
QY 752 NTPSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVQLTQMLPYGCLLDHVRENRGL 811
Db 175 ATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLTQMLPYGCLLDYIREHKDNI 234
QY 812 GSODLLNWCQIAKMSYLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEV 871
Db 235 GSOYLLNWCQIAKMSYLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEV 294
QY 872 HADGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIPLEK 931
Db 295 HAEGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGSKPYDGIPISEISSVLEK 354
QY 932 GERLPQPPICITIDVYIMVVKCWMIDSECRPRFRELVSFMRMARDPQRFVIO-NEDLGP 990
Db 355 GERLPQPPICITIDVYIMVVKCWMIDADSRPKFRELIAEFSKMARDPPEYLVIOGDERMHL 414
QY 991 ASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCCPDAPGAGGMVYHRHRSSTRSRG 1050
Db 415 PPTOSKFTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 454
QY 1051 GDLTLGLEPSEEAAPRSP-----APSEGAGSDVDFDGLGMAAKGLQSLTHDPPSPLO 1105
Db 455 -----SRTPLLSLSATSNNATNCID-----RNGQHPVREDSFVQ 492
QY 1106 YSEDPVLPSET--DGYVAPLTCSPQPEYVQDVRPQPPSPREGPLPAARPAATLER 1163
Db 493 YSSDPTGNFLEESIDGFL-----PAPEYVQ--LMPKKPS----- 526
QY 1164 AKTLPCKNGVVKDVF-----AFGAVENPEYL 1191
Db 527 -----TAMVQNIYNNISLTAISKLPMDSRYSQNSHSTAVDNPEYL 566

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag v-erbA, v-erb-B protein.
GN GAG, V-ERB-A, V-ERB-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Bruskin A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RA "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B.";

```

```

RL Oncogene 5:15-24(1990).
CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -.
DR EMBL; X52211; CAA36459.1; JOINED.
DR HSSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Steroid_receptor.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHOMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000001; Euk_pkinase; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; TyrKc; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4B95CE CRC64;

Query Match 23.7%; Score 1610.5; DB 15; Length 962;
Best Local Similarity 50.9%; Pred. No. 1e-112;
Matches 352; Conservative 67; Mismatches 149; Indels 123; Gaps 17;

QY 541 VEECRVLOGLPRE-YVNAR-HCLP-----CHPSCQ 568
DB 354 IEKQESYLLAFEHVINYRKGNIPHFWSKLLMKVADLRMIGAYHASRFLHMKVECPTELS 413
QY 569 PONGSVTCFGEADOCVAHYKDPFCVAPCPGKPDLSYMPYKVPDEEGACQCPPI 628
DB 414 PQE-----VGP--DHCCKCAHFIDGPHCVKACPCAGVLGENDTL-VWKYADANAVCQLCHP 465
QY 629 NCTHSCVDLDKGCPEQRASPLTSIQYIKANSKF-----IGITELKRRQOKIRKYTMW 682
DB 466 NCTRCCKGFLGECGP---NGSKTPSIAAGVVGGLCLLVVGLGILYLRHRHIVRKRTL 522
QY 683 RLLOETELVEPLTPSGAMPNQAOMRIKTELKRVKVLGSGAFGVYKGIWIPDGENVKI 742
DB 523 RLLOERELVEPLTPSGEAPNQAHLRIKTEFEKVKVGLGFGAFGVYKGLWIPGEKVTI 582
QY 743 PVAIKVLRENTSPKANKEILDEAYVMAGVGSPPYSRLIGICLTSTVQLVTQMLPYGCLLD 802
DB 583 PVAIKELRENTSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQMLPYGCLLD 642
QY 803 HVRENRRGLSGDQLLNCWQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLAR 862
DB 643 YIREHKONIGSQYLLNWCQIAKGNVLEERHVRDLAARNVLKTPQHKITDFGLAK 702
QY 863 LLDIDETEHADGGKVPKWMALLESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPA 922
DB 703 QLGADEKYEAEKGVPIKWMALLESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPA 762
QY 923 REIPDLLEKGRLLPQPPCTTIDVYIMVWKWIMISECRPRFRELVSFSPMRDQPRV 982
DB 763 SEISSVLEKGRLLPQPPCTTIDVYIMVWKWIMISECRPRFRELVSFSPMRDQPRV 822
QY 983 IQ-NEDLGPASPLDSTFYSRLEDDDDGLVDAREYLVPOQGFPCDPAPAGAGMWHRH 1041
DB 823 IQGDERMHLPSPTUSKFTVRLMEEDMEDIVDAEYLVPHQGF----- 866
QY 1042 RSSSTRSGGDLTLGLFSEEEAPRSLAPSEGAGSDVFDGLGMAKGLQSLPHTDPS 1101
DB 867 NSPST-----SRTPLLSLSATS-----NSATKCIDRNGGH--- 898

```

```

QY 1102 PLQRYSEDDTVELPSETDGYVAPLTCSPQPEYVNPQDVRPQPSPREGPLPAARAGAT- 1160
DB 899 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNIYVYISLTA 937
QY 1161 LERAKTLPSPKNGVVKDVFAFGCAVENPEYL 1191
DB 938 ISKLPMDSRVQN-----SHSTAVDNPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (T934) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
NCBI_TaxID=11861;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CAA30024.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DC8CCA0F8AF4 CRC64;

Query Match 23.6%; Score 1602; DB 15; Length 545;
Best Local Similarity 54.1%; Pred. No. 2e-112;
Matches 339; Conservative 64; Mismatches 128; Indels 96; Gaps 14;

QY 578 GPEADQCACAHYKDPFCVAPCPGKPDLSYMPYKVPDEEGACQCPINCTHSCVDL 637
DB 1 GP--DHCCKCAHFIDGPHCVKACPCAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGP 57
QY 638 DDKGCPAEQRASPLTSIQYIKANSKF-----IGITELKRRQOKIRKYTMRLLOETELV 691
DB 58 GLEGCP---NGSKTPSIAAGVVGGLCLLVVGLGILYLRHRHIVRKRTLRLLOERELV 114
QY 692 EPLTPSGAMPNQAOMRIKTELKRVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLE 751
DB 115 EPLTPSGEAPNQAHLRIKTEFEKVKVGLGFGAFGVYKGLWIPGEKVTIPVAIKELRE 174
QY 752 NTSKANKEILDEAYVMAGVGSPPYSRLIGICLTSTVQLVTQMLPYGCLLDHVRNCRGL 811
DB 175 ATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQMLPYGCLLDYIREHKONI 234
QY 812 GSQDLLNCWQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEY 871
DB 235 GSQYLLNWCQIAKGNVLEERHVRDLAARNVLKTPQHKITDFGLAKQLGNADEKEY 294
QY 872 HADGGKVPKWMALLESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEK 931
DB 295 HAEGGKVPKWMALLESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPASEISSVLEK 354
QY 932 GERLPQPPCTTIDVYIMVWKWIMISECRPRFRELVSFSPMRDQPRVVIQ-NEDLGP 990
DB 355 GERLPQPPCTTIDVYIMVWKWIMISECRPRFRELVSFSPMRDQPRVVIQOGDERMHL 414

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.5911 Seconds
(without alignments)
4891.279 Million cell updates/sec

Title: SEQ4-632-652-12
Perfect score: 6796
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6619	97.4	1255	1	P06426 homo sapien
2	5857	86.2	1257	1	ERB2_HUMAN
3	5833.5	85.8	1254	1	ERB2_MESAU
4	3113	45.8	1210	1	EGFR_HUMAN
5	3083	45.4	1210	1	EGFR_MOUSE
6	2925.5	43.0	1308	1	ERB4_HUMAN
7	2907	42.8	1308	1	ERB4_RAT
8	2651.5	39.0	1167	1	XMRK_XIPMA
9	2407.5	35.4	1342	1	ERB3_HUMAN
10	2320.5	34.1	1339	1	ERB3_RAT
11	1944	28.6	1426	1	EGFR_DROME
12	1706.5	25.1	634	1	ERBB_ALV
13	1660	24.4	604	1	ERBB_AVIER
14	1587	23.4	540	1	ERBB_AVIEU
15	1552	22.8	703	1	EGFR_CHICK
16	1282	18.9	1323	1	LT23_CABEL
17	1142.5	16.8	245	1	ERB2_MOUSE
18	711	10.5	1363	1	ILPR_BRALA
19	690	10.2	1300	1	IRR_MOUSE
20	679	10.0	1382	1	INSR_HUMAN
21	679	10.0	1607	1	MTPR_LYMST
22	678.5	10.0	1300	1	IRR_CAVPO
23	678	10.0	1297	1	IRR_HUMAN
24	678	10.0	1383	1	INSR_RAT
25	677.5	10.0	1372	1	INSR_MOUSE
26	673	9.9	1477	1	HPK7_HYDAT
27	632	9.3	1367	1	IGIR_HUMAN
28	618	9.1	1373	1	IGIR_MOUSE
29	615	9.0	1390	1	INSR_AEDAE
30	614.5	9.0	1370	1	IGIR_RAT
31	612	9.0	2146	1	INSR_DROME
32	588	8.7	757	1	HT16_HYDAT
33	581.5	8.6	1053	1	FAK1_CHICK

34	576.5	8.5	984	1	EPB1_CHICK
35	576	8.5	1114	1	RET_HUMAN
36	574.5	8.5	984	1	EPB1_RAT
37	569	8.4	1068	1	FAK1_XENLA
38	568.5	8.4	984	1	EPB1_HUMAN
39	568	8.4	987	1	EPB4_HUMAN
40	568	8.4	1052	1	FAK1_MOUSE
41	568	8.4	1055	1	FAK1_RAT
42	567	8.3	902	1	EPBB_XENLA
43	565.5	8.3	746	1	ABL_MLVAB
44	565	8.3	1052	1	FAK1_HUMAN
45	557.5	8.2	977	1	EPA2_MOUSE

ALIGNMENTS

RESULT 1	ERB2_HUMAN	STANDARD;	PRT; 1255 AA.
ID	AC	P04626;	
DT	13-AUG-1987	(Rel. 05, Created)	
DT	13-AUG-1987	(Rel. 05, Last sequence update)	
DT	15-JUN-2002	(Rel. 41, Last annotation update)	
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)		
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell		
DE	surface receptor HER2) (MLN 19).		
GN	ERBB2 OR HER2 OR NGL OR NEU.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
[1]	SEQUENCE FROM N.A.		
RP	MEDLINE=86118663; PubMed=3003577;		
RX	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,		
RA	Saito T., Toyoshima K.;		
RT	"Similarity of protein encoded by the human c-erbB-2 gene to		
RT	epidermal growth factor receptor.";		
RL	Nature 319:230-234(1986).		
[2]	SEQUENCE FROM N.A.		
RP	MEDLINE=86070181; PubMed=2999974;		
RX	Cousens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,		
RA	McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,		
RA	Francke U., Levinson A., Ullrich A.;		
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor		
RT	shares chromosomal location with neu oncogene.";		
RL	Science 230:1132-1139(1985).		
[3]	SEQUENCE OF 737-1031 FROM N.A.		
RP	MEDLINE=86016729; PubMed=2995967;		
RX	Semba K., Kamata N., Toyoshima K., Yamamoto T.;		
RA	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the		
RT	c-erbB-1/epidermal growth factor-receptor gene and is amplified in a		
RT	human salivary gland adenocarcinoma.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).		
[4]	VARIANTS VAL-654 AND VAL-655.		
RP	MEDLINE=93194196; PubMed=8095488;		
RX	Ehsani A., Low J., Wallace R.B., Wu A.M.;		
RA	"Characterization of a new allele of the human ERBB2 gene by allele-		
RT	specific competition hybridization.";		
RL	Genomics 15:426-429(1993).		
CC	-!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,		
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A		
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-		
CC	ALPHA AND AMPHIREGULIN.		
CC	-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein		
CC	tyrosine phosphate.		
CC	-!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS		
CC	-(POTENTIAL).		
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.		

FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 86.2%; Score 5857; DB 1; Length 1257;
Best Local Similarity 85.8%; Pred. No. 2.6e-301;
Matches 1083; Conservative 51; Mismatches 110; Indels 18; Gaps 4;

QY 1 MELAAALCRWGLLALLPPGAASCTOVCTGDMKRLRPASPETHLDMLRHLVGGCQVQGNL 60
DB 1 MELAAWCRWGLLALLPPGAGTQVCTGDMKRLRPASPETHLDMLRHLVGGCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEQVGYVLIAHNQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
DB 61 ELTYVSPANASLSFLQDIOEQVGYVLIAHNQVQVPLQRLRIVRGTLQFEDKYALAVLDNR 120
QY 121 DPLNNTTPVT-GASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILKWDIFHKKNQL 179
DB 121 DPQDNVAASPTPGTPEGLRELQRLSLTEILKGGVLIQORNPOLCYQDMVLWKVDVFRKNQL 180
QY 180 ALTLIDNRRGRACHPCSPCKMGSRGWESSEDCOSLRTVTCAGGCARCKPLPTDCCHEQ 239
DB 181 APVDIDNRRGRACHPCSPCKMGSRGWESSEDCOSLRTVTCAGGCARCKPLPTDCCHEQ 240
QY 240 CAAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFFGASCVTAC 299
DB 241 CAAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMHNPGRYTFFGASCVTTC 300
QY 300 PYNLYSTDVSGSCTLVCPHLNHOEVTAEQGTORCEKSPCARVCYGLGMOYIKANSKFIGI 359
DB 301 PYNLYSTEVSGSCTLVCPNPNQEVTAEDGTORCEKSPCARVCYGLGMEHLRGARAITSD 360
QY 360 TELEFACCKIFGSLAFLPESFDGDPASNTAPLOEQLOVFTELEETIGLYTSAPDLSL 419
DB 361 NVQEFDCCKIFGSLAFLPESFDGDPSSGATAPLEQLOVFTELEETIGLYTSAPDLSL 420
QY 420 PDLVSFQNLQVIRGRILHNGAYSITLQGLGISWGLSLRELGSGLALIHNNTHLCFVHT 479
DB 421 RDLVSFQNLRIIRGRILHNGAYSITLQGLGTHSLGLSLRELGSGLALIHNNHLCFVHT 480
QY 480 VPNDOLFERNPHOALLHTANPEDE-CVGEGLACHQLCARGHCWCPGTQCVNCSQFLRGQ 538
DB 481 VPNDOLFERNPHOALLHSGNRPEDLCVSSGLVNSLCAHGCWCPGTQCVNCSHFHURGO 540
QY 539 ECVEECRVLOGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPECVA 598
DB 541 ECVEECRVWGLPREYVYVSDKRLPCHPECPQNSSETCFGEADQCAACHYKDXSSCVA 600
QY 599 RCPGSGVKPDLSYMPIWPKPDEEGACQPCPNCTHSCVDLDDKGCAPASORASPLTSIOYIK 658
DB 601 RCPGSGVKPDLSYMPIWPKPDEEGICQPCPNCTHSCVDLDERGCPAQRASPVTFI---- 656
QY 659 ANSKFIGI-----TELKEROOKIRKTYMRRLLQETELVEPLTBSGAMPNQOMR 707
DB 657 -IATVGVGLFLILVVVVGILIKRRKRIKRYMRRLLQETELVEPLTBSGAMPNQOMR 715
QY 708 ILKETELRKVKVILGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKETILDYAV 767
DB 716 ILKETELRKVKVILGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKETILDYAV 775
QY 768 MAGVSGPYVSRLLIGICLTSTVQLVTLMPYGLLDHVRNRRGRIGSODLLNMCQIAKGM 827

DB 776 MAGVSGPYVSRLLIGICLTSTVQLVTLMPYGLLDHVRNRRGRIGSODLLNMCQIAKGM 835
QY 828 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALES 887
DB 836 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALES 895
QY 888 ILARRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYM 947
DB 896 ILARRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYM 955
QY 948 IMVKCMWIDSECRPRFRELVSFSESRMARDPQRVFIQNEIDLGPASPLDSTFYRSLLEDD 1007
DB 956 IMVKCMWIDSECRPRFRELVSFSESRMARDPQRVFIQNEIDLGPSSPMDSTFYRSLLEDD 1015
QY 1008 MGLVDAEYLVPOQGFPCDDPAPGAGGMVHRRSRSSSTRSGGDLTLGLSPSEEEAPRS 1067
DB 1016 MGLVDAEYLVPOQGFPCDDPAPGAGGMVHRRSRSSSTRSGGDLTLGLSPSEEGPRS 1075
QY 1068 PLAPSEGAGSDVFDGDLGMAAGKLOSLPHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1127
DB 1076 PLAPSEGAGSDVFDGDLGMAAGKLOSLPHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1135
QY 1128 SPOPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKOVFAFGGAVEN 1187
DB 1136 SPOPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKOVFAFGGAVEN 1195
QY 1188 PEYLTOGGGAAPHPHPPAPSPAFDNLVYWDQPPPERGAPPTFKGTPTAENPEYLGVDV 1247
DB 1196 PEYLVPREGTASPHPSAFDNLVYWDQPPPERGAPPTFKGTPTAENPEYLGVDV 1255
QY 1248 PV 1249
DB 1256 PV 1257

RESULT 3
ERB2_MESAU STANDARD; PRT; 1254 AA.
ID ERB2_MESAU Q0553;
LT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERBB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura Y., Uehijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Iehikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the ENBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@sb-sib.ch).

DR	EMBL; D16295; BAA03801.1; .
DR	HSP3; P11362; IFGK.
DR	InterPro; IPR000494; EGFR_L_domain.
DR	InterPro; IPR000719; Euk_pkinase.
DR	InterPro; IPR002174; Furin-like.
DR	InterPro; IPR001245; Tyr_pkinase.
DR	InterPro; IPR004019; YLP_motif.
DR	Pfam; PF00069; pkinase; 1.
DR	Pfam; PF00757; Furin-like; 1.
DR	Pfam; PF01030; Recep_L_domain; 2.
DR	Pfam; PF02757; YLP; 2.
DR	ProDom; PD000001; Euk_pkinase; 1.
DR	SMART; SM00261; FU; 3.
DR	SMART; SM00219; TyxKC; 1.
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW	Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW	Proto-oncogene; Disease mutation.
FT	SIGNAL 1 21 POTENTIAL.
FT	CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT	DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 653 675 POTENTIAL.
FT	DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 158 368 CYS-RICH.
FT	DOMAIN 472 644 CYS-RICH.
FT	DOMAIN 720 987 PROTEIN KINASE.
FT	NP_BIND 726 734 ATP (BY SIMILARITY).
FT	BINDING 753 753 ATP (BY SIMILARITY).
FT	ACT_SITE 845 845 BY SIMILARITY.
FT	DLSLFD 195 204 BY SIMILARITY.
FT	DLSLFD 199 212 BY SIMILARITY.
FT	DLSLFD 236 244 BY SIMILARITY.
FT	DLSLFD 240 252 BY SIMILARITY.
FT	DLSLFD 255 264 BY SIMILARITY.
FT	DLSLFD 268 295 BY SIMILARITY.
FT	DLSLFD 299 311 BY SIMILARITY.
FT	DLSLFD 315 331 BY SIMILARITY.
FT	DLSLFD 334 338 BY SIMILARITY.
FT	DLSLFD 511 520 BY SIMILARITY.
FT	DLSLFD 515 528 BY SIMILARITY.
FT	DLSLFD 531 540 BY SIMILARITY.
FT	DLSLFD 544 560 BY SIMILARITY.
FT	DLSLFD 563 576 BY SIMILARITY.
FT	DLSLFD 567 584 BY SIMILARITY.
FT	DLSLFD 587 596 BY SIMILARITY.
FT	DLSLFD 600 623 BY SIMILARITY.
FT	DLSLFD 626 634 BY SIMILARITY.
FT	DLSLFD 630 642 BY SIMILARITY.
FT	MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARIANT 658 658 V -> E (IN ONCOGENIC NEU).
FT	VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
FT	VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
SO	SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

```

Oy 1070 APSEGAGSDVFDGLGMAAGKLOSLPHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSF 1129
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1076 APSEGAGSDVEGELGKATGKPOSISPRDLSPLQRYSEDTPLPTETDGYVAPLACSP 1135
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 1130 QPEYVQPDVVRPQPPSPREGPLPAARAGATLERAKTLSPKNGVVDVFAFGAVENPE 1189
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1136 QPEYVQPEVRPQPLTPEGLPPVRPAGATLERAKTLSPKNGVVDVFTFGAVENPE 1195
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 1190 YLTGCGNAPOPHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1196 YLVRGGSASQPH-PPALCPAFDNLVYWDQDPPSERGSPNTFEGTPTAENPEYLGLDVVP 1254

RESULT 4
EGFR_HUMAN
ID EGFR_HUMAN STANDARD; PRT: 1210 AA.
AC P05333; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;
AC O06688; Q9B252; Q9H2C9; Q9GZX1; Q9H3C9;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
DE protein-tyrosine kinase ErbB-1).
GN EGFR OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Varden Y., Libermann T.A., Schlessinger J., Downward J.,
RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells."
RL Nature 309:418-425(1984).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=95382957; PubMed=7654368;
RA Ilekis J.V., Stark B.C., Scoccia B.;
RT "Possible role of variant RNA transcripts in the regulation of
RT epidermal growth factor receptor expression in human placenta."
RL Mol. Reprod. Dev. 41:149-156(1995).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=97078686; PubMed=8918811;
RA Reiter J.L., Maible N.J.;
RT "A 1.8 kb alternative transcript from the human epidermal growth
RT factor receptor gene encodes a truncated form of the receptor."
RL Nucleic Acids Res. 24:4050-4056(1996).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=97256547; PubMed=9103388;
RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
RT "Expression of a truncated epidermal growth factor receptor-like
RT protein (TEGFR) in ovarian cancer."
RL Gynecol. Oncol. 65:36-41(1997).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RC TISSUE=Placenta;
RX MEDLINE=21100872; PubMed=11161793;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative EGFR transcripts encoding truncated receptor
RT isoforms."
RL Genomics 71:1-20(2001).

```

```

RN [6]
RP SEQUENCE OF 575-687 FROM N.A.
RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maible N.J.;
RT "Human and mouse alternative EGFR transcripts encoding only the
RT extracellular domain of the receptor."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; PubMed=6326261;
RA Lin C.R., Chen W.S., Krueger W., Stolarek L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT "Expression cloning of human EGF receptor complementary DNA: Gene
RT amplification and three related messenger RNA products in A431
RT cells."
RL Science 224:843-848(1984).
RN [8]
RP SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245835; PubMed=6330563;
RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RT "Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAs overproduced in A431 carcinoma cells."
RL Nature 309:806-810(1984).
RN [9]
RP SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; PubMed=6093780;
RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
RA O'Malley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RT receptor cDNA from human A431 carcinoma cells."
RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN [10]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=88217333; PubMed=3329716;
RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.D.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription."
RL Oncogene Res. 1:375-396(1987).
RN [11]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; PubMed=1988448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis."
RL J. Biol. Chem. 266:1746-1753(1991).
RN [12]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; PubMed=2991899;
RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene."
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN [13]
RP SEQUENCE OF 540.
RA Kohda D.;
RL Submitted (SEP-1997) to the SWISS-PROT data bank.
RN [14]
RP RECEPTOR ACTIVITY.
RX MEDLINE=84191554; PubMed=6325948;
RA Mroczkowski B., Mosig G., Cohen S.;
RT "ATP-stimulated interaction between epidermal growth factor receptor
RT and supercoiled DNA."
RL Nature 309:270-273(1984).
RN [15]
RP PHOSPHORYLATION.
RX MEDLINE=89278137; PubMed=2543678;
RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
RT "All autophosphorylation sites of epidermal growth factor (EGF)

```

receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor.",
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN [16]
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RL "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts.",
 RL Growth Factors 13:121-132(1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 ASN-603.
 RX MEDLINE=20198209; PubMed=1073168;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RL "Characterization of the N-oligosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor.",
 RL J. Biochem. 127:65-72(2000).
 RN [18]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Odaka M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
 RL "Disulfide bond structure of human epidermal growth factor receptor.",
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN [19]
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RL "Receptors for epidermal growth factor and other polypeptide
 RT mitogens.",
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, Gp30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X00588; CAA25240.1; -;
 DR EMBL; U95089; AAB53063.1; -;
 DR EMBL; U48722; AAC50802.1; -;
 DR EMBL; U48723; AAC50804.1; -;
 DR EMBL; U48724; AAC50796.1; -;
 DR EMBL; U48725; AAC50797.1; -;
 DR EMBL; U48726; AAC50798.1; -;

Query Match 45.8%; Score 3113; DB 1; Length 1210;
 Best Local Similarity 49.5%; Pred. No. 8.6e-157;
 Matches 624; Conservative 173; Mismatches 360; Indels 104; Gaps 21;

QY 11 LLLALLPPGAA--STOVCTGTDMLRLPASPETHLDMLRLHYOGCVVQVGNLELYLPTN 68
 DB 14 LLAALCPASRALEEKKVCOQTSNKLTLQGTGFEDHFLSLQRMFNWCEVVLGNLEITYVORN 73
 QY 69 ASLSFLQDIOEVQGYVLIHNRVQVPLQRLRVRGTLQFEDNYALAVLDNGDPLNNTTP 128
 DB 74 YDLSFLKTIQEVAGYVLIALTVERIPLENLQIRGNMYENSVALAVLSNYD----- 126
 QY 129 VTGASPGGLRELQRLSLTEILKGVLIORNPOLCYQDTILWKDIFHKQNQLALTLIDTNR 188
 DB 127 ---ANKTGLKELPMRLQELHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDQFNH 183
 QY 189 SRACHPCSPMKGSRGSESDCOSLTRTVACGGCA-RCKGGLPPTDCCHEQCAAGCTGP 247
 DB 184 LGSCKQCDPCPNCSGNGWAGEENCQKLTIKIAQOCSGRGCRGKSPSCCHNQCAAGCTGP 243
 QY 248 KHSDDLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCTVACPNYIISTD 307
 DB 244 RESDCLVCRKFRDEATCKDTPPLMLNPTTYQMDVNPPEGKYSFGATCVKKCPNRYVTD 303
 QY 308 VGSCTLVCPHNOEVTAEQDQCEKSKFCARVCYGLGHQYIKANSKFTGITELE-PAG 366
 DB 304 HGSQVACGADSYEM-BEDGVRKCKKCEGCRKVCNGIGIGEPEK-DSLSINATNIKHPKN 361
 QY 367 CKKIFGSLAFPESEGDGPASNTAPLOPEOLQVFPETLEEITGYLYISAWPDSLPLDSVFQ 426
 DB 362 CTSISGDLHLPLVAFRGDSFTHTPDPQBLDIKTVKBITGFLUIQAMPENRTDLHAPE 421
 QY 427 NLQVIRGRIILHNGAYSILTLQGLGTSWGLRSLRSLRELGLALIHNTLHLCFVHTVPDOLF 486
 DB 422 NLEIIRGRTKHGQFSLAVVSLNITSLSLASLKEISDGDVIISGNKNLCYANTINMKKLF 481
 QY 487 RNPHQALLHTANRPEDECVEGEGELACHOLCARHCWGPGPTQCVNCSOFLRQECVEECRV 546
 DB 482 GTSQGTKIISNRGENSCKATQVCHALCSPEGCGPEPRDCVSRNVSREGCDKCNL 541
 QY 547 LQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVAPCPSGVKP 606
 DB 542 LESEPREFVENSECIIQCHPECLPQAMNITCTGRGPDNCTIQCAHYIDGPHCVKTCFAGVNG 601
 QY 607 DLSYMPIWKFPEBEGACQPCPINCTHSCVDLDDKGPAPORASPLTSTQYIKANSKF--- 663
 DB 602 ENNTL-VWKYADAGHVCHLCHPNTCYCTGPGLEGCTNGPKIPSIATGMVALLLVV 660
 QY 664 -IGITELRRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKETELRKVKVLGS 722
 DB 661 ALGTGLFMRRRHVRKTRRLRLQERELVEPLTPSGEAPNQALLRLKETEFKKIKVLGS 720
 QY 723 GAFCTVYKGIWDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVSPYVSRLLGI 782
 DB 721 GAFGTVYKGLWIPGEKVKIPVAIKELREATSPKANKEIIDEAYVMASVDNPHVCRLLGI 780
 QY 783 CLTSTVOLVTQLMPYGCLLDHRNRLGRSLQDLNLCMOIAKGMVSLDVLVHRDLAA 842
 DB 781 CLTSTVOLITQLMPFGCLLDYVREHKDNIGSVLLNLCVQIAKGMVSLDVLVHRDLAA 840
 QY 843 RNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALESIILRRRTHOSDVMYSY 902
 DB 841 RNVLKTPQHVKITDFGLAKLLGAEEKEHYAEGKVPKWMALESIILHRYTHOSDVMYSY 900
 QY 903 GVTVMELMTGAPYDGIIPAREIPDLLEKGERLPQPPCTIDVTVMVWKHMIIDSECRPR 962
 DB 901 GVTVMELMTGSKPYDGIIPASEISSILEKGERLPQPPCTIDVTVMVWKHMIIDSECRPR 960
 QY 963 FRELVSFESRMARDPQRFVVIQ-NEDLGASPLDSTFYRSLLEDDEDDGLVDAEYLVQ 1021
 DB 961 FRELIIFESKARDPQRYLVIQDERMHLSPDTSNRYALMDEEDMDVDADEYLIPIQ 1020
 QY 1022 QGFFCPDPAPGAGMVHRRSSSTRSGGDLTLGLPEPSEEAAPRSPPLASEGAGSDVFD 1081
 DB 1021 QGFF-----SSPSTSRTPLLSLSATS--N 1044

```

OY 1082 GDLGMAAKQLSLTHDPSPLOQYSEDPTVPLPSET--DGYVAPLTCSPQPVVNPQDV 1139
DB 1045 STVACIDRNLQSCPIKEDSFLOQYSDPTFGALTEDSIDDTFL-----PVPYINQ-SV 1097
OY 1140 RPOPPSPREGPLPAAPAGATLERAKTLPSPKNGVVKVFAFGGAVENPEYL-TPQGGAA 1198
DB 1098 PKRPAGSVQNPVYHNPQPLN-----APSRDPHYQD--PHSTAVGNPEYLVNTVQ---- 1143
OY 1199 PQPHPPAPSPAFDNLYYNDQ-----DP-----PERCAPPTSTFKGTPTAENPEY 1242
DB 1144 -----PTCVNSTFSDPAHWAQKSHQISLDNPDYQDFFPPEAKPNKGFKGS-TAENAEY 1197

OY 1243 L 1243
DB 1198 L 1198

RESULT 5
EGFR_MOUSE STANDARD; PRT; 1210 AA.
AC Q01279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
RT (bek/KGFR) gene.";
RL Oncogene 7:1957-1962(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RT in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Hibbs M.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RT receptor tyrosine kinase.";
RL Genes Dev. 8:399-413(1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91212866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
RT "Comparison of EGF receptor sequences as a guide to study the ligand
RT binding site.";
RL Oncogene 6:673-676(1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eisinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.

```

```

CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X78987; CAA55587.1; -
CC EMBL; U03425; AAA17899.1; -
CC EMBL; X59698; CAA42219.1; -
CC EMBL; L06864; AAA53029.1; -
CC EMBL; Z12608; CAA78249.1; -
CC HSSP; P11362; 1FGK.
CC MGD; MGI:95294; Egfr.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1
CC Pfam; PF00757; Furin-like; 1
CC Pfam; PF01030; Recep_L_domain; 2
CC ProDom; PD000001; Euk_pkinase; 1
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyrKC; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
CC SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1210 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 25 647 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 648 670 POTENTIAL.
FT DOMAIN 671 1210 CYTOPLASMIC (POTENTIAL).
FT REPEAT 75 300 APPROXIMATE.
FT REPEAT 390 600 APPROXIMATE.
FT DOMAIN 1028 1071 SER-RICH.
FT DOMAIN 714 981 PROTEIN KINASE.
FT NP_BIND 720 728 ATP (BY SIMILARITY).
FT BINDING 747 747 ATP (BY SIMILARITY).
FT ACT_SITE 839 839 BY SIMILARITY.
FT DISULFID 190 199 BY SIMILARITY.
FT DISULFID 194 207 BY SIMILARITY.
FT DISULFID 215 223 BY SIMILARITY.
FT DISULFID 219 231 BY SIMILARITY.
FT DISULFID 232 240 BY SIMILARITY.
FT DISULFID 236 248 BY SIMILARITY.
FT DISULFID 251 260 BY SIMILARITY.
FT DISULFID 264 291 BY SIMILARITY.
FT DISULFID 295 307 BY SIMILARITY.
FT DISULFID 311 326 BY SIMILARITY.
FT DISULFID 329 333 BY SIMILARITY.
FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.

```

FT	DISULFID	620	628	BY SIMILARITY.
FT	DISULFID	624	636	BY SIMILARITY.
FT	MOD_RES	680		PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT	MOD_RES	1092		PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1110		PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1172		PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1197		PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT				(BY SIMILARITY).
FT	CARBOHYD	128	128	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	175	175	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	196	196	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	352	352	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	413	413	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	444	444	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	528	528	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	568	568	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	603	603	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT	19	19	C -> S (IN REF. 2).
FT	CONFLICT	539	539	C -> W (IN REF. 5).
FT	CONFLICT	991	991	L -> F (IN REF. 4).
FT	CONFLICT	1116	1117	HP -> DR (IN REF. 6).
SQ	SEQUENCE	1210 AA;	134853 MW;	690E20D46DF2D2F5 CRC64;
Query Match				
Best Local Similarity 45.4%; Score 3083; DB 1; Length 1210;				
Matches 630; Conservative 164; Mismatches 362; Indels 116; Gaps 25;				
QY	11	LLALLPPGAA--STQVCTGTDMKRLPASPEETHLMRLHVGQVQVGNLELTVLPTN	68	
DB	14	LTALCAAGALEEKVCGTSENRLTQGTDFDHFLSLQRMVNCVVLGNLEITVYQRN	73	
QY	69	ASLSFLQDIQEVQGYVLIHAHQVQVLPQRLRIVRGTLQFEDNYALAVLDNGDPLNNTTP	128	
DB	74	YDLSFLKTIQEVAGYVLIHAHQVQVLPQRLRIVRGTLQFEDNYALAVLDNGDPLNNTTP	128	
QY	129	VTGASPGGLRELQSLRLTEILKGGVLIQRNPOLCVQDTILWKDI----	184	
DB	125	YGTNRRTGLRELPMNLQILIGAVRFSNNPLCNMDTIQWRDVIQNVFMNSNMDL----	180	
QY	185	DNRRSRACHPCSPMKSGSCWSESSEDCOSLRTVCAGGCA--RCKGPLTDCCHQCAAG	243	
DB	181	-OSHPSSCPKDPSPGSCWGGGNCOKLTKIACAQCSHRCRGRSPSDCCNQCAAG	239	
QY	244	CTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVCATCPVNY	303	
DB	240	CTGPRESCLVCQKQFQDEATCKDTCPLMLYNTTYQMDVNFEGKYSFGATCVKCKPRNY	299	
QY	304	LSTDVGSCTLVCPHNOEVAEDGTQRCCKSPCARVCYGLGMQVYKANSKFIGITELE	363	
DB	300	VYTDHGSCVRACGPDYEV--EEDGIRKCKKCDGCRKVCNGIGIGEFK--DTLSINATNIK	357	
QY	364	-PAGCKKIFGSLAFLEPGFDGPPASNTAPLOEQVQFETLEITGYLVISAWPDSLPL	422	
DB	358	HPKYCTALSGDLHILPFAFGKDSFTPLDPRELEILTKVEITGFLLIQAWPNDTDL	417	
QY	423	SVFQNLQVIRGRLHNGAYSLTLOGLSWGLSLRELIGSLALIHNNHLFCFVHTVPW	482	
DB	418	HAFENLEIIRGRTKQHGQFSLAVVGLNITSLGLSLKEISDGDVVISGNRLCYANTINW	477	
QY	483	DOLFENPHQALLHTANRDECECVGEGSLACHQLCARGHCWGPPTQCVNCSQFLRQECVE	542	
DB	478	KKLFOTPNOKTKIMNRAEKDKAVNHNVCNPLCSSEGCWGPPEPRDCVSNVSRGECVE	537	
QY	543	ECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDRPPFCVARGPS	602	
DB	538	KCNILEGEPREVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYDGHCHVKTCPA	597	
QY	603	GVPKPLSYMPIWKFPEEGACQCPINCTHSCVDLDDKGPAPQASPLTSTIQY--IKANS	661	
DB	598	GIMGENNTL--VMKYADANNVCHLCHANCTYGCAGPLQGCCEWMPSPGPKIPSIATGVGL	656	
QY	662	KFI-----GITELKRQKIRKYTWRRLLQETLEVEPLTPSGAMPNQAMRILKETELRK	716	

DB	657	LFIVVVALGIGLFMRRRHIVRKRTRRLRLQERLVEPLTPSGEAPNQAHLRIKETEKK	716	
QY	717	VKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIIDEAVVMAGVGSYV	776	
DB	717	IKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIIDEAVVMASVDNPHV	776	
QY	777	SRLLGICLTSTVOLVTLQMPYGLCLLDHVRNRRGRGLSDQLLNMCMQIAKMSYLEDVRLV	836	
DB	777	CRLLGICLTSTVOLVTLQMPYGLCLLDYVREHKDNIGSQVLLNWCVCQIAKGMYLEDRRLV	836	
QY	837	HRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADCGKVPKIMMALESILRRPFTHQ	896	
DB	837	HRDLAARNVVKTPQHVKITDFGLAKLLGAEKEKEYHABGGKVPKIMMALESILHRYTHQ	896	
QY	897	SDVMSYGVTVWELMTFGAKPYDGIIPARBIPLDLLEKGERLPQPPICITIDVYIMVVKCWMID	956	
DB	897	SDVMSYGVTVWELMTFGSKPYDGIIPASDISLLEKGERLPQPPICITIDVYIMVVKCWMID	956	
QY	957	SECRPRFRELSEFSRMARDPQRFVVIQ--NEDLGPASPLDSTFYRSLLEDDDDMDLVDAE	1015	
DB	957	ADSRPKFRELILEFSKWARDPQRYLVIOGDERMHLPSPTDSNFYRALMDEEDMEDVVDAD	1016	
QY	1016	EYLVPOQGFCCPDPAAGACGMVHHRRSSSTRSGGDLTLGLEPSEEBAPRSPPLASEGA	1075	
DB	1017	EYLIPQOGFF-----NSPST-----SRTPLLSLSA	1042	
QY	1076	GSDFDGLDGMCAAGLQSLPHTDPSPLQRYSEDTPVLPSET--DGYVAPLTCSPQPEY	1133	
DB	1043	TSN-----NSTVACIWRNGSCRKVEDAFQRYSDPTGAVTEDNIDDAFL-----PVPEY	1092	
QY	1134	VNQDVRPQPPSPREGPLPAAPAGATLERAKTSLPGKNGVVKDVPFAGGAVENPEYL--T	1192	
DB	1093	VNQ--SVPKPEAGSVQNPVYHNPQLHP-----APGRDLHYQN--PHSNAVGNPEYLT	1141	
QY	1193	PGGAAPOPHPPAPSPADNLYWDO-----DP-----PERGAPSTFTKGTPT	1236	
DB	1142	AO-----PTCLSSGFNSPALMTQKSHQMSLDNPDYQODFFPKETKPKNGIFKG--PT	1191	
DB	1237	AENPEYGLDVP	1248	
DB	1192	AENAEYLRVAPP	1203	
RESULT 6				
ID	ERR4_HUMAN	STANDARD;	PRT;	1308 AA.
AC	Q15303;			
DT	15-DEC-1998	(Rel. 37, Created)		
DT	15-DEC-1998	(Rel. 37, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)			
DE	(p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).			
GN	ERBB4 OR HER4.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM JM-A).			
RC	TISSUE=Breast carcinoma;			
RX	MEDLINE=93189574; PubMed=8383326;			
RA	Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,			
RA	Foy L., Neubauer M.G., Shoyab M.,			
RT	"Ligand-specific activation of HER4/p180erbB4, a fourth member of the			
RT	epidermal growth factor receptor family,"			
RL	Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM JM-A AND JM-B).			
RX	TISSUE=Fetal brain;			
RX	MEDLINE=97476287; PubMed=9334263;			
RA	Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,			
RA	Klagsbrun M.;			

RT "A novel juxtamembrane domain isoform of HER4/Erbb4. Isoform-specific
RT tissue distribution and differential processing in response to
RT phorbol ester".

RL J. Biol. Chem. 272:26761-26768(1997).

CC -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC

CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
CC

CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC

CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC

CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;
CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.

CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
CC KIDNEY. IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
CC LUNG, SALIVARY GLAND, AND PANCREAS.

CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC

CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; L07869; AAB59446.1; -;
CC HSSP; P11362; 1FGK.
CC

DR Genew; HGNC:3432; ERBB4.
DR MIM; 600543; -;
DR

DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR

DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR

DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR

DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR

DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR

DR SMART; SM00261; FU; 4.
DR SMART; SM00219; TyrKc; 1.
DR

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR

DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW

KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW alternative splicing.
FT

FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT

FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT

FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 186 334 CYS-RICH.
FT

FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE.
FT

FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT

FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT

FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT

FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT

FT	DISULFID	234	246	BY SIMILARITY.
FT	DISULFID	249	258	BY SIMILARITY.
FT	DISULFID	262	289	BY SIMILARITY.
FT	DISULFID	293	304	BY SIMILARITY.
FT	DISULFID	308	323	BY SIMILARITY.
FT	DISULFID	326	330	BY SIMILARITY.
FT	DISULFID	503	512	BY SIMILARITY.
FT	DISULFID	507	520	BY SIMILARITY.
FT	DISULFID	523	532	BY SIMILARITY.
FT	DISULFID	536	552	BY SIMILARITY.
FT	DISULFID	555	569	BY SIMILARITY.
FT	DISULFID	559	577	BY SIMILARITY.
FT	DISULFID	580	589	BY SIMILARITY.
FT	DISULFID	593	614	BY SIMILARITY.
FT	DISULFID	617	625	BY SIMILARITY.
FT	DISULFID	621	633	BY SIMILARITY.
FT	MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	174	174	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	181	181	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	253	253	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	358	358	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	410	410	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	473	473	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	495	495	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	548	548	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	620	620	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	626	648	NGPTSHDCIYPMKTHSTLPQHA -> IGSSIEDICGLMD (IN ISOFORM JM-B)
FT	SEQUENCE	1308 AA;	146807 MW;	5E4AE0985D88761 CRC64;

Query Match 43.0%; Score 2925.5; DB 1; Length 1308;
Best Local Similarity 45.2%; Pred.No. 7.1e-147;
Matches 608; Conservative 175; Mismatches 385; Indels 177; Gaps 28;

QY	9	WGILLALLPGAA----	STQVCTGTDMLRLPASPETHLDMLRLHYGCGVOVGNLELT	64
DB	8	WYVSLVAAGTVQPSDSQSCAGTENKLSLSLEQQYRALRKYYENCEVVMGNLEITS	67	
QY	65	LPTNASLSFLQDIQEVQGVVLIAHNOVROVPLQRLRI	VRGTOLFDENYALAVLNDGDP	124
DB	68	IEHRDLSFLRSVREVTGVVLVALNOFRYLPLENRI	IRGTCLYEDRYALAFINVRKOG	127
QY	125	NTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPOLCYQDTILKWKDIFHKNQLALT	LI	184
DB	128	NF-----GLQELGLKLNTEILNGVYVDQNKFLCYADTIHMQDIVRNWPNSNLT	LV	178
QY	185	DTNRSRACHPCSPMCKGSCWGESSEDCQSLTRTVCAAGC-ARCKGSLPTDCCHEQCAAG	243	
DB	179	STNGSSGCRCHKSCGTG-RCWGPTEHNHCQTLTRTVCAEQCDRCYGPVYSDCCHRECAGG	237	
QY	244	CTGPKHSDCLACHFNHSGICELHCPALVYNTDTEFSPNPEGRVTFGASCYVACP	PYNY	303
DB	238	CSGPKDTCFACFNFDNSGACVTCQPTVFYNPPTTFQLEHNFNAKVFACFKKCPHNF		297
QY	304	LSTDVGSCTLVCLPHNQEVTAEDGTQRCCKSPCARVCYGLGMQVIMKANSKEFIGITE	LE	363
DB	298	V-VDSSCVRACPSKMEV-EENGIMCKPCTDPCPKACDGICTGSLMSAQTVDDSSNIDK		355
QY	364	FAGCKIFGSLAFLPESFGDDPASNTAPLOPQOLQVFETLEBRTGYLYISAWPDSLP	DL	423
DB	356	FINCTKINGNLIFLVTGIHGDYPYNAIEADPEKLVNFRVREITGFLNQSPWPNMTD	FS	415
QY	424	VFQNLQVIRGRIHNCAYSLTLOGLGISWLRSLRELGLALHHNTHCFVHTVP	WD	483
DB	416	VFSNLVTIGRVLYSGLSLILKQOQISLQFOSLKEISAGNIYITDNSNLCYYHTIN	MT	475
QY	484	QLFRNPHQALLHTANRPEDECVCGEGLACHQLCARGHCWGPQTQCVCNCQFQ	LRGQCEV	543

Db 476 TLFTINQIRIVIRNRKAENCTAEGMVCNHLCSGDCGMPGPDQCLSCRRFSRGRICIES 535
Qy 544 CRVLOGLPREYVNAHCLPCHPEQOP-ONGSVTCFGEADOCVACAHYKOPPFVCARCP5 602
Db 536 CNLYDGEFREPENGSICVECDQCEKMDGLTCHGPDNCTKCSHFPGDPCNVCVKCPD 595
Qy 603 GVKPDLSPMTWKFPDEGACQPCPINCHSCVDLDDKGC-----PABQASPL 651
Db 596 GLOGANSF--IPKYADPDRECHPCPNCTQCNGPTSHDCIYYPWKGHSTLPQHAR-TPL 652
Qy 652 TSIQVIKANSXFI-----GIT--ELKRROQKIRKYTMRLLOETELVEPLTPSGAMPNQAQ 705
Db 653 IAGVIL--GGILFIVLGLTFAVYVRRKS IKKRALRRFL-ETELVEPLTPSGTAPNQAQ 709
Qy 706 MRILKTELKRVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLENTSPKANKEILDEA 765
Db 710 LRILKTELKRVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLENTSPKANKEILDEA 769
Qy 766 YMGVGVSPYVRLGICLTSTVQLVLTOLMPYGCILLDVRNRRGLSGQDLLNWCMIQAK 825
Db 770 LIMASMDPHVLRLGLVCLSPITQILVTQMLPHGCLLEYVHEKDNIGSQLLNWCMIQAK 829
Qy 826 GMSYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIMMAL 885
Db 830 GMYLEERLVRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIMMAL 889
Qy 886 ESILRRRTHQSDWSYGVTVVHMLTFGAKPYDGIPIAREIPDLLEKGERLPOPPCTIDV 945
Db 890 ECIHYRKHQSDWSYGVTVVHMLTFGAKPYDGIPIAREIPDLLEKGERLPOPPCTIDV 949
Qy 946 YMWKCMWIDSECRPRFRELSEFMRMARDPQRFVVIQNE--LGPASPLDSTFYRSLLE 1004
Db 950 YMWKCMWIDSECRPRFRELSEFMRMARDPQRFVVIQNE--LGPASPLDSTFYRSLLE 1009
Qy 1005 DDDMGDLVDAEYLVPOQGFCCPDAPAGAGGMVHRRSSSTRSGGGDTLTGLPESEEA 1064
Db 1010 EEDLEDMDAEYLVLP-QAFNIPPP-----IVTSRAIDSNRS-----EIGHSPPPAYT 1057
Qy 1065 PRS-----PLAP-SEGAGSDVFDGLGMAKGLQSL 1095
Db 1058 PMSGNOFVYRDGGFAAEGQSVSPYRAPTSPTPEAPVAQGATAEIFDDSCCNGLRKPVP 1117
Qy 1096 PTHDPSPLQRYSDPTVPLPS-----ETDGVVAPLTCSPQPEYVNPQDVRPQPSPRE 1148
Db 1118 HVQEDSSTORYSADPTVFAPERSPRGELDEGYWTPMRDKQEYLPV-----E 1168
Qy 1149 GPLPAAPAGATLERAKTLSPGKNGVVKVQVFAFGAVENPEYLTPOGGAAPQHPHPPA-- 1206
Db 1169 NPFVSR-----KNGDLQ-----ALDNPETHNASNG-----PPKAE 1200
Qy 1207 -----ESPADFNLYYNDQDPPERGA--PPSTFK 1232
Db 1201 EYVNEPLYLNTFANTLGAEYLVKNILSMPEKAKAFDNPDYMNHSLPPRSTLQHPDYLQ 1260
Qy 1233 GTPT-----AENPEYL 1243
Db 1261 EYSTKYFYKQNGRIRPIVAENPEYL 1285
RESULT 7
ERB4_RAT
ID_ERB4_RAT STANDARD; PRT; 1308 AA.
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_taxid=10116;
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS. NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF041838; AAD08999.1; -.
CC EMBL; U52531; AAC53051.1; -.
CC HSP; P11362; IFGK
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR00719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC InterPro; IPR004019; YLP motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC Pfam; PF02757; YLP; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00281; FU; 4.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00111; PROTEIN KINASE DOM; 1.
CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25

FT	CHAIN	26	1308		RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT	DOMAIN	26	651		EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	652	675		POTENTIAL.
FT	DOMAIN	676	1308		CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	186	334		CYS-RICH.
FT	DOMAIN	496	633		CYS-RICH.
FT	DOMAIN	718	985		PROTEIN KINASE.
FT	NP BIND	724	732		ATP (BY SIMILARITY).
FT	BINDING	751	751		ATP (BY SIMILARITY).
FT	ACT SITE	843	843		BY SIMILARITY.
FT	DISULFID	189	197		BY SIMILARITY.
FT	DISULFID	193	205		BY SIMILARITY.
FT	DISULFID	213	221		BY SIMILARITY.
FT	DISULFID	217	229		BY SIMILARITY.
FT	DISULFID	230	238		BY SIMILARITY.
FT	DISULFID	234	246		BY SIMILARITY.
FT	DISULFID	249	258		BY SIMILARITY.
FT	DISULFID	262	289		BY SIMILARITY.
FT	DISULFID	293	304		BY SIMILARITY.
FT	DISULFID	308	323		BY SIMILARITY.
FT	DISULFID	326	330		BY SIMILARITY.
FT	DISULFID	503	512		BY SIMILARITY.
FT	DISULFID	507	520		BY SIMILARITY.
FT	DISULFID	523	532		BY SIMILARITY.
FT	DISULFID	536	552		BY SIMILARITY.
FT	DISULFID	555	569		BY SIMILARITY.
FT	DISULFID	559	577		BY SIMILARITY.
FT	DISULFID	580	589		BY SIMILARITY.
FT	DISULFID	593	614		BY SIMILARITY.
FT	DISULFID	617	625		BY SIMILARITY.
FT	DISULFID	621	633		BY SIMILARITY.
FT	MOD RES	1162	1182		PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD RES	1188	1188		PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD RES	1258	1258		PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD RES	1284	1284		PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	138	138		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	174	174		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	253	253		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	358	358		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	410	410		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	473	473		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	495	495		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	548	548		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	576	576		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	620	620		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT	1062	1062		S -> N (IN REF. 3).
FT	CONFLICT	1080	1082		PYT -> SYR (IN REF. 3).
SQ	SEQUENCE	1308 AA;	146957 MW;	D944BB096A08B41 CRC64;	
Query Match 42.8%; Score 2907; DB 1; Length 1308;					
Best Local Similarity 44.9%; Pred. No. 6.7e-146;					
Matches 605; Conservative 182; Mismatches 395; Indels 164; Gaps 27;					
Qy	1	MELA-ALCRWGLLL--ALLPGAASVQCTGDMKRLRPASPETHLDMLRLHYGCGVQV 57			
Db	1	MKLATGLVWGSLLVAARTVQPSASQSVACAGTENKLSLSLEQQYRALRYKYYENCEVVM 60			
Qy	58	GNLELTLYPTNASLSFLQDIOEVOGYVLIHANOVRQVFLORLIRVRGTOLFEDNVALAVL 117			
Db	61	GNLEITSIEHRNDSLFLRSIREVTGYVLVALNQRYLPLENLRIRGTLYEDYALAI 120			
Qy	118	DNGDPLNNTTPTVGASPGRLRELQRLRSITELKGGVLIQORNPOLCYQDTILWKDIFHKNN 177			
Db	121	LNRYKDGNF-----GLQELGLKNLEITLNGGVYVDQNKFLCYADTIHWQDIVRNPW 171			
Qy	178	QLATLTDNRSRACHPCSPMKSGRSGESSEDCQSILTRTVACAGGC-ARCKGLPLTDCC 236			
Db	172	PSNMWLVSTIGSSGCGRCHKSCG-RCMGPTENHCQTLTRTVCAEQDGRCYGYPVSDCC 230			
Qy	237	HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFFESMPNPEGRYTFGASCV 296			
Db	231	HRECAGCGSGPKDTCFACMFNDNSGACVTCQPTQTFVYNPTTFQLEHNFNAKYTGAFCV 290			

Qy	297	TACPYNVLTSDVSGSCTLVCPLHNOEVTAEADGTORCEKSKPCARVCYGLGMOYIKANSKF 356			
Db	291	KKCHNFV-VDSSCVCRACSSKMEV--EENGIKMCKPCTDICKACDGIPTGSLMSAQTV 348			
Qy	357	IGITELEFAGCKKIFGSLAFILPESFDGDPASNTAPLOEQLOVETLEETITGLYLYISAWP 416			
Db	349	DSSNIDKFINCTKINGNLIFLVTGIHGDPVYNAIDAIDPEKLVNFRVREITGFLNIQTWP 408			
Qy	417	DSLPLSVFQNLQVIRGRILHNGAYSITLQGLGISWGLRSLRGLSGSLAIHNNHLCF 476			
Db	409	PNMTDFSVFSLNLTIGRVLYSGLSLILKQGGITSQFQSKLSISAGNIYITDNSLCY 468			
Qy	477	VHTVPQLFRNPQHALLHTANRPEDECVEGLACHQLCARGHCWGPCTQVCNCSQFLR 536			
Db	469	YHTINWTLTSTVNRIVIRNDRRAENCTAGMVCNHLCSNDGCMGPGDQCLSCRRFSR 528			
Qy	537	GOECVEBCRVLOGLPREYVNAHCLPCHPEQCP-ONGSVTCFGEADOCVCAHYKDPFF 595			
Db	529	GKICIESNLVDGFEFBFENGSI CVECDSSQCEKMEGLLTCHGPGPDNCTKCSHFKDGPN 588			
Qy	596	CVARCPGVKPDLSYMPITWKFPPDEEGACQCPINCTHSCVDLDDKGC-----PA 644			
Db	589	CVEKCPDVLQANSF--IFKYADODRECHPCHPCTQCGNPTSHDCIYYPWTGHSTLUPQ 646			
Qy	645	EQRASPLTISQYIKANSKFIGI-----TELKRRQOKIRKYTMRLRLLQETELVEPLTPSG 698			
Db	647	HAR--TPLIAGVI--GGFLVILVINALTFVYVYRKSIIKKKALARRFL--ETELVEPLTPSG 702			
Qy	699	AMPNQAQRILKETELRKVKVGLSGAFGVYVGIWI PDGENVKIPIVAIKVLENTSPKAN 758			
Db	703	TAPNQAQLRILKETELRKVKVGLSGAFGVYVGIWVPEGETVKIPIVAIKILNETTGPKAN 762			
Qy	759	KEILDEAVVWAGVSPVUSRLGLTCLSTVOLVTLQMPYGLCLLDHVRNRCGLSODLLN 818			
Db	763	VEFDEALIMASVDPHLVLGLVCLSTPTQLVTLQMPHGLCLLYVHEHKONIGSOLLN 822			
Qy	819	WCMTAKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYIADGGKV 878			
Db	823	WCVOIAGMWMYLEERLVRDLAARNVLKSPNHVKITDFGLARLLLEGDEKEYNADGGM 882			
Qy	879	PIKMALESILRRRFTHOSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPOP 938			
Db	883	PIKMALECIHYRKFTHQSDVMSYGVTVWELMTFGKPYDGIPTREIPDLLEKGERLPOP 942			
Qy	939	PICITDVMIMVKCMWIDSECRPRFRELVSFSSWARDPQRFVUIONED-LGPASPLDST 997			
Db	943	PICITDVIYVWVKCMWIDADSRPKFELAAEFSSWARDPQRYLVIOGDDRMKLPSPNDSK 1002			
Qy	998	FYRSLLEDMDGDLVDAEYLVPOQGFPCDP-----APCAG 1034			
Db	1003	FFQNLLDEEDLEDMWDAEYLVQ-ANENIPPIYSTRIDSNRSEICHSPPPAYTPMSG 1061			
Qy	1035	GMVHRRSSSTRSGGDLTLGLPESEEA PRSPLAPSEGAGSDVFDGDLGMAAGKLGOS 1094			
Db	1062	SQFYVQGGFATQGG---MPMPYTATTSTIPEADVA--QGATAEMFDDSCCNGTLRKPVV 1116			
Qy	1095	LPHTDPSPLORYSEDPVPLPS-----ETDGVVAPLTCSPQPEYVNPQDVRPQPSPR 1147			
Db	1117	PHVQEDSSTQRYADPTVFAPERNPRAELDEEGYMTPMHDKPOEYLNPIV----- 1167			
Qy	1148	EGPLPAARPAAGATLERAKTLSPGKNGVYVQVFAFGGAVENPEYILTPOCGAAPQPHPPA- 1206			
Db	1168	ENPPVSR-----KNGDLQ-----ALDNPETHSASSG-----PPKAE 1199			
Qy	1207	-----PSPAFDNLYYWDQPPPERGA--PPSTF 1231			
Db	1200	DEYVNEPLYLNTFTNALGNAEYMKNSLLSVPEKAKKAFDNPYWNHSLPPRSTLQHPDYL 1259			
Qy	1232	KGITP-----AENPEYL 1243			
Db	1260	QESTKYFYKONGRIRPIVAENPEYL 1285			

```

RESULT 8
XMRK_XIPMA
ID_XMRK_XIPMA STANDARD; PRT; 1167 AA.
AC P13388;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN XMRK OR TU.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphia; Acanthopterygii; Percomorpha; Atherinomorphia;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8083;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=90015140; PubMed=2797166;
RX Wittbrodt J., Adam D., Malitschek B., Maueler W., Raulf F.,
RA Telling A., Robertson S.M., Scharl M.;
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT inducing Tu locus in Xiphophorus.";
RL Nature 341:415-421(1989).
RN [2]
REVISION TO 515.
RA Scharl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X16891; CAA34770.2; -
CC FIR; S06142; S06142.
CC HSP; P11362; LFPG.
CC InterPro: IPR000494; EGFR_L_domain.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR002290; Ser_thr_pkinase.
CC InterPro: IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 5.
CC SMART; SM00220; S_TKc; 1.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
CC PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT SIGNAL 1 25
FT CHAIN 26 1167
FT DOMAIN 26 642
FT TRANSMEM 643 665
FT DOMAIN 666 1167
FT DOMAIN 710 977
FT NP_BIND 716 724
FT BINDING 743 743
FT ACT_SITE 835 835

```

```

FT DISULFID 195 204
FT DISULFID 199 212
FT DISULFID 220 228
FT DISULFID 224 236
FT DISULFID 237 245
FT DISULFID 241 253
FT DISULFID 256 265
FT DISULFID 269 296
FT DISULFID 300 311
FT DISULFID 315 330
FT DISULFID 333 337
FT DISULFID 504 513
FT DISULFID 508 521
FT DISULFID 524 533
FT DISULFID 537 553
FT DISULFID 556 569
FT DISULFID 560 577
FT DISULFID 593 615
FT DISULFID 618 626
FT DISULFID 622 634
FT CARBOHYD 114 114
FT CARBOHYD 144 144
FT CARBOHYD 201 201
FT CARBOHYD 356 356
FT CARBOHYD 365 365
FT CARBOHYD 398 398
FT CARBOHYD 417 417
FT CARBOHYD 501 501
FT CARBOHYD 576 576
FT CARBOHYD 621 621
FT SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;
Query Match 39.0%; Score 2651.5; DB 1; Length 1167;
Best Local Similarity 45.1%; Pred. No. 1.7e-132;
Matches 571; Conservative 160; Mismatches 393; Indels 143; Gaps 29;
QY 4 AALCRWGLLLALPPGAASL---QVCTGDMKRLPASPETHLDMRLHLYQCGQVVOGN 59
DB 8 AALQ---LLLVLSISCCSTDPDRKVCQGTSNMTW---LDNHYLKKMKMYSGCNVULEN 62
QY 60 LEITYLPTNASLFLQDIQEVGVLIHNOVQVPLQRLRIVRGTQLFEDNALAVLDN 119
DB 63 LEITYQENQDLSFLQSIQEVGVLIHNEVSTIPLVNLRLIRGQNLVEGNTLLVMSN 122
QY 120 GDPLANTTEVTGASPGGLRELQLRLSLEILKGGVLIQRPOLCYQDTILWKDIFHKGNQL 179
DB 123 YQK-NPSSP--DYVQVGLQQLQSLNLEILSGGVKVSHPNLLCNVETINWWDIVDKTSNP 179
QY 180 ALTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSILTRTVCAAGC-ARCKGPLPTDCCHE 238
DB 180 TNNLIPHAFERQCKCDHGCVNGSCWAPGPGHCQKFTKLLCARQCNRRCRGPKPIDCCNE 239
QY 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTA 298
DB 240 HCAGGCTGPRATDCLACRDFNDGDKTCPPPKIYDIVSHQVVDNPKIYTFGAACVKE 299
QY 299 CPNYLSTDVSGCTLVCPPLHNOEVTAEQDGTQRCCKSKPCARVCYGLGMQYIKANSKFIG 358
DB 300 CFSNVVTE-GACVRSKCSAGMLEVD-ENGKRSCKPCDGVCPKVCDGIGISL-SNTIAVN 356
QY 359 ITFL-EFAGCKIFGSLAFLEPESFGDGPASNTAPLOEQLOVFPETLEEITGYLISAWPD 417
DB 357 STNIRSFNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLNLTITVKEITGYLVIMWPE 416
QY 418 SLPDLISVFQNLQVIRGRILHNGAYS-LTLQGLIGISWLGRLSRLRELQSLALIHNTLCF 476
DB 417 NMTLSLVFQNLLEIRGRTFSGFVQVQVRLHQLWGLRLSRLSEVAGNVLKNTLQLRY 476
QY 477 VHTVPWDLFRNPQALLHTANRPEDECVCVGEGLACHOLCARHCWGPGGTQCVNCSQFLR 536
DB 477 ANTINWRLFRSEDSOIEYDART-----ENQTCNNECSGCGWGGPTM-VSLHVDNR 529
QY 537 GQECVECEKRVLQGLPREVYNARHCLPCHPEQPGNSVTCFGPEADQCACAHYKDPFPC 596

```

```
Db 530 GGRCVASCNLLQGEPREAQVDRGCVQCHQECVLQVDTSLTCYGFGRANCSKSAHFQDGPQC 589
Qy 597 VARPSPGVKPDLSYMPIWKPFPDEGACQPCINCHTSCVDLDDKGGPAE-----QRASPL 651
Db 590 IPRCPHGILGDGDTL-IMKYADKMGQCPQCHQCTGCGSGPGLSGCRGDIIVSHSLAVGL 648
Qy 652 TSIOYIKANSKFGITELKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRLIKE 711
Db 649 VSGLLTIVIVALLIVLLRRRRK-RKRTIRCLLOEKELVEPLTPSGAQNQAFRLIKE 707
Qy 712 TELRKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGV 771
Db 708 TEFKDRVLGSGAGFTYKGLMNPDPGENIRIPVAIKVLRENTSPKANKEILDEAYVMAGV 767
Qy 772 GSPVSELGICITSTVOLVTLMPYCLLDHYVNRGRGLSGDOLLNMCQIAKMSYLE 831
Db 768 DHPHVCRLGICITSAVOLVTLMPYCLLDHYVNRGRGLSGDOLLNMCQIAKMSYLE 827
Qy 832 DVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRR 891
Db 828 ERHLVHRDLAARNVLKSPNHVKITDFGLSKLLTADEKEYQADGGKVPKMALESILQW 887
Qy 892 RFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVK 951
Db 888 TYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKETASVLENGERLPQPICTIEVYMIILK 947
Qy 952 CWMIDSECRFRFELVSEFMRARDPQFVVQINEDLGPASPLDSTFYRSLLEDMDGDL 1011
Db 948 CWMIDSECRFRFELVSEFMRARDPQFVVQINEDLGPASPLDSTFYRSLLEDMDGDL 1002
Qy 1012 VDAEYLVPQGGFFCPDPAPGAGVWHRHRSSTRSGGGDLTLGLPESEEAAPRSLAP 1071
Db 1003 VDAEYLVPYKRI-----NRQGS-----EFCIP 1025
Qy 1072 SEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSDPVP-PLPSTDGYVAPLTCSPQ 1130
Db 1026 PTGH-----PVRENSITLRNISDPTQNALEKDLGDH----- 1056
Qy 1131 PEYVNPQDVRPQP-----PSPRE-----GSLP-AARPAGATLERAKTLPNGKGVVKD 1177
Db 1057 -EYVNPQSTSRSLSDIYNPNVEDLTDGMPVSLSSQEAETNFSREYLNNTQNSL--- 1112
Qy 1178 VFAFGAVENPEYLTPOGGAAPQHPPPAFSPFDNLYWDQDPPPERGAPPSTFKGTPTA 1237
Db 1113 PLYSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTNGMFLPAA 1150
Qy 1238 ENPEYLG 1244
Db 1151 ENLEYLG 1157
```

RESULT 9

```
ERB3_HUMAN STANDARD; PRT; 1342 AA.
ID ERB3_HUMAN
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
```

```
Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
[2]
SEQUENCE FROM N.A.
RP MEDLINE=90311312; PubMed=2164210;
RA Plowman G.D., Whittier G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RP Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
[3]
SEQUENCE FROM N.A. (SHORT FORM).
RP TISSUE=Placenta;
RC MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M29366; AAA35790.1; -
CC EMBL; M34309; AAA35979.1; -
CC EMBL; S61953; AAB26935.1; -
CC PIR; A36223; A36223.
CC HSSP; P11362; 1FGK.
CC Genew; HGNC:3431; ERBB3.
CC MIM; 190151; -
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_Pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_Pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC ProDom; PD000001; Euk_Pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
CC PROSITE; PS00109; PROTEIN KINASE TYR; FALSE_NEG.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transfrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Alternative splicing.
CC SIGNAL 1 19 POTENTIAL.
CC CHAIN 20 1342 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
CC DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 644 664 POTENTIAL.
CC DOMAIN 665 1342 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 709 966 PROTEIN KINASE.
CC NP_BIND 715 723 ATP (BY SIMILARITY).
CC
```

```
FT BINDING 742 742 ATP (BY SIMILARITY).
FT ACT_SITE 834 834 BY SIMILARITY.
FT DISULFID 186 194 BY SIMILARITY.
FT DISULFID 190 202 BY SIMILARITY.
FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 141 183 SC -> GQPMVPSGLTPQPAQDWYLLDDPRLLTSSAK
VPVTLAAV (IN SHORT ISOFORM).
FT VARSPLIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;

Query Match 35.4%; Score 2407.5; DB 1; Length 1342;
Best Local Similarity 40.5%; Pred. No. 1.4e-119;
Matches 531; Conservative 189; Mismatches 455; Indels 137; Gaps 33;

Qy 10 GLLIALLPPGAA--STQVCTGTDMLRLPASPTHLDMLRHLYQGCQVQVQGNLEIYLTPT 67
Db 11 GLLFSLARGSEVGNQAQVCPGTLNGLSVTGAENQYQTLKLYERCEVVMGNLEIVLTGH 70

Qy 68 NASLSFLQDIOEVGYVLIANQVQVPLQRLIRVGTQLFDNEDYALAVLDNGDPLNNTT 127
Db 71 NADLSFLQWIREVTGYVLIANNEFSTLPLNLRVVRGTQVYDGKFAIPVM-----LNVT 125

Qy 128 PVTGASPGLEQLRLSITLTKGVLGIQNPOLCYOCTILWKDIFHKNQALALTLDTN 187
Db 126 ----NSSHALRQLRLTQLTEILSGVVIKNDKLCHMDTIDWRDIVRDRD---AEIVVKD 178

Qy 188 RSRACHPCSPMKGRWCWGESSEDCQLSRLTRVCAGGC-ARCKGPLPTDCHEOCAGCTG 246
Db 179 NGRSCPPECHVKCG-RCMWPGSEDCQLTKTKICAPQCNGCHGFCGNPNQCCHECAGCGSG 237

Qy 247 PKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPDEGRYTFGASCTVACPNYLS 306
Db 238 PQDTDCFACRHFNDSGACVPRCPQPLVYNNKLTFOLEPNHTKYQYGGVGVASCAPHNFV-V 296

Qy 307 DVSGSCTLVCLHNOEVTAEQDQRCCKSPCARCYGLGQYIKANSKF--IGITILE- 363
Db 297 DQTSQVACRPPDKMEVD-KNGLKWCFCGGLCPKACEGTG-----SGSRFQTVDSNIDG 350

Qy 364 FAGCKKIFGSLAFIPESDFGASNTAPLOEQLOVFETILEITGYLYISAWPDSLPLDS 423
Db 351 FVNCTKILGNLDFLITGLNGDPWHKIPALDPEKLNVTREITGYLYNQSWPPHMFNS 410
```

RESULT 10

ERB3_RAT

ID_ERB3_RAT

STANDARD; PRT; 1339 AA.

AC Q62799; Q62955;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)

DE (c-erbB3).

GN ERBB3.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RA "Cloning of the rat ErbB3 cDNA and characterization of the
RT recombinant protein";
RL Gene 165:279-284(1995).
RN [2]
RP REVISIONS TO 85; 513 AND 565.
RA Hellyer N.J., Koland J.G.;
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RA "Expression of neuroligins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U29339; AAC28498.2; -;
DR EMBL; U52530; AAC53050.1; -;
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGF_R domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR PRINTS; PR001030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transpherase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19
FT CHAIN 20 1339
FT DOMAIN 20 643
FT DOMAIN 644 662
FT TRANSMEM 663 1339
FT DOMAIN 183 259
FT DOMAIN 707 964
FT NP_BIND 713 721
FT BINDING 740 740
FT ACT_SITE 832 832
FT DISULFID 186 194

FT DISULFID 190 202 BY SIMILARITY.
FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1028 1028 L -> P (IN REF. 3).
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBDF1E CRC64;
Query Match 34.1%; Score 2320.5; DB 1; Length 1339;
Best Local Similarity 40.5%; Pred. No. 5.4e-115;
Matches 520; Conservative 170; Mismatches 430; Indels 165; Gaps 35;
QY 3 LAALCRWGLLLALPPGAA---STQVCTGTOMKRLPASPETHLDMLRLHYOGCQVVOGN 59
DB 7 LQVLC---FLLSLARGSEMNSQAVCPGTGLNGLSVTGDADNOVQTLKLYEKEVVMGN 62
QY 60 LELTYLPNLSLFLQDIQEVQGVLYIAHNOVQVLPQRLRIVRQTQFEDNYALAVLDN 119
DB 63 LEIVLTGHNADLSFLQWIREVTGVLYVANNEFSLVPLNLRVVRGTQVYDGFALFVM-- 120
QY 120 GDPINNTTPVTGASPGGLRELQLRSLTEILAGGVLIQNPOLCYQDTILMKDIFHKNQL 179
DB 121 ---LNYNT---NSSHALRQLKFTQLTEILSGGVYIEKNDKLMHMDTIDWRDVRVR-- 170
QY 180 ALTLIDTNRSRACHPCSPMKGSRGSGWSSSDCSQSLTRTVCAAGC-ARCKGPLPTDCCH 238
DB 171 GAELVVKXNGANGCPCPCHEVCKG-KCWGPGDDCQLTKTKICAPQCNRCFGPNPQCCHD 229
QY 239 QCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTA 298
DB 230 ECAGGCGSPQDTDCFACRRFNDSGACVPRCPPELVYNKLTQLEPNPHTKYQGVGCVAS 289
QY 299 CPVNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGL--GMQYIKANSKF 356
DB 290 CPHNFV-VDQTFVCRACPPDQNEVD-KHGLKMCPEPCGGLCPKACEGTSGSGRYQTVDSSN 347
QY 357 IGITELSEFAGCKIFGSLAFLPESPDGDPASNTAPLOPEQLQVFTLEITGLYLISAWP 416
DB 348 ID----GFVNCKTKILGNLDFLITGLNVDPWHKIPALDEKLVNFTVREITGLYNISQWP 403
QY 417 DSLPDLVSFQNLQVIRGRLHNGAYS-LTLOGLGISWGLSLRSLRELGLALIHNTHLIC 475
DB 404 PHMFNFVSFSLNLTIGRSLYNRGFSLLIMKNLNVTSLSFRSLKEISAGRVVISANQQLC 463
QY 476 FVHTVPWDQLPFRNPQALLHTA-NRPDECEVGEGLACHQLCARGHCWCGPPTOCVNCQSF 534
DB 464 YHSLNWTLLRGPSEERLDIKYDRPLGECIAEGKVCDPCLCSSGCGWGPQGCLSCRNY 523
QY 535 LRQGECEVEECRLVQLGLPREYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDP 594


```
Db 524 SREGVCVTHCNFLOQEPREFVHEAQCFSHPCELPMEGTSTCNGSGSDACARCAFRDGP 583
Qy 595 FCVAPCPGVRDLSYMPIWKPPDEGACQPCPINCSTHSCVDLDDKGPAPORASPLTSI 654
Db 584 HCVNCSCHPILG--AKGPIYKYPDQAQNECRPCHECTGCGNPELQDCLGQ--AEVLMASK 639
Qy 655 QY-----IKANSKFTGITELKRRQOKIR-KYTMRRLLQETLVEPLTPSGAMPNQAQ 705
Db 640 PHLVATVTVGLAVILMILGGSFLYWRGRIQKRAMRRYLERGESIEPLDPS-EKANKVL 698
Qy 706 MRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEA 765
Db 699 ARIFKETELRKVLGSGVGVGVHKGIMIPGESIKIPVCIKVIEDKSGRQSFQAVTDM 758
Qy 766 YVAGVGSPYVSRLLIGLICLTSTVLQVLTQMPYGLCLDHDVRENRGLGSDLLNWCQIAK 825
Db 759 LAVGSLDHAHIVRLGLCPGSSQLVQVPLPLGSLLDHVKQHRRTLGPOLLNMGVQIAK 818
Qy 826 GMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETVHADGGKVPKKNMAL 885
Db 819 GMYLDEHSVHRDLARNVLMKSPSQVADFGVADLLPPDKQLLHSEAKTPKKNMAL 878
Qy 886 ESILRRRTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDV 945
Db 879 ESIHFGKYTHQSDVMSYGVYVWELMTFGAEPYAGRLAEIPDLLEKGERLAQPQICTIDV 938
Qy 946 YMIWKCMWIDSECPRELVSERFMRDQRFVIVQNEQDLGASPLDSTFYRSLLED 1005
Db 939 YVMIWKCMWIDENIRPTFKELANETFRWARDPPRYLVTKRAS-GPGTTP--PAAEPSVLTT 995
Qy 1006 DMGDLVDAEEVLVPOQGFCCPDPAFGAGGMVHRRSSRSGGDLTLGLEPSEE--- 1062
Db 996 KEL-----QEALEFEL-----DLDLDEAEEGEA 1021
Qy 1063 -----EAPRSLPAPSEG-----AGSDVFDGDLGMAAKGLQSLPT 1097
Db 1022 TSLGSALSPLTCTLRPRGSQLSPSSGYMPMNQSSLGCEACLSAVLGGREGFRSPISL 1081
Qy 1098 HDPSPLOYRSEDPVPLPSETDGVY-----APL-----TC-----SPOPE-----YVNPQ 1137
Db 1082 H-PIPRGR-----PASESECHVTGSEAELOEKVSCRSRSPRPRGDSAYHSQR 1133
Qy 1138 DVROPSPPREPG-----LPAARPAGATLERAKTLP-SGKGVV-----KQVF 1179
Db 1134 HSLTPTVPLSPGLLEEDGNGYVMPDTHLRGASSREGTSLVGLSSVLGTREEDED--- 1191
Qy 1180 AFGGAVENPEYLTPOGGAAPQPHPP 1204
Db 1192 -----EEVEYNNRRKRGSP-PRPP 1209
```

RESULT 11

```
EGFR_DROME
ID EGFR_DROME STANDARD; PRT: 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (Torpado protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
that several genetically defined classes of alleles cluster in
```

```
RT subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RN [2]
RP REVISIONS.
RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RX STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Bannkoch C., Milos G.L.G.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinartz K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupek M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.
```



```
Db 543 RPAATQKEPEQKWNENLRADLCEKNGTICSDQNCDCGAGTDCQCTCKNFNGTC 602
QY 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPFCVARC 600
Db 603 IADCGVISNAYK--FDNRCKICHPECR-----TCNGAGADHCQECVHVRDQHCVCSEC 654
QY 601 P-----SGVK-----PDL 608
Db 655 PKNKYNDRCVCRECHATCGCTGPKDITIGIGACTTCNLAIINNDATVVRCKLLKDDKCPD- 713
QY 609 SYMPIWKF--PDEEGACQP-----CPI-----NCTH-----632
Db 714 GY--FWYVHPQEQSLKPLAGRAVCRKCHPLCELCNTGYGHEQVCSKTHYKRRREQCBT 771
QY 633 -----SC-----VDLDDKG-----641
Db 772 ECPADHYDEQRECFQRHPECNCGTGPAGDDCKSCRNPKFLFDANETGYPVNSTWPNCTFS 831
QY 642 -CPAEQR-----ASPLTSIQY---IKANSKFI-----GITEL 669
Db 832 KCPLMRHVNYOYTAIGPYCAASPRSSKITANLDVNMIFITGAVLVPTICILCVVTYI 891
QY 670 KRRQOKIRKYT--MRLQLQETELVEPLTPSGAMPNQAMRILKETELRKVKVLGSGAGCT 727
Db 892 CRQOKAKKETVMTMALSGCEDSEPLRPSNIGANLKLRIKVDKELRGVGLHGAFGR 951
QY 728 VYKGWIIPDGENVKIIPAIVKURENTSPKANKEILDEAYVMAGVSPYVRLGLICLTST 787
Db 952 VYKGVVPEGENVKIIPAIVKELLKSTGAESSEFLREAYIMASEHVNLLKLLAVCMSSQ 1011
QY 788 VOLVTQMPYGCGLDVRNRCGLCSODLLNWCQIAKMSVLELDVRLVHRDLAARNVLV 847
Db 1012 MMLITQLMPLGCLLDYVRNRNDKIGSKALLNWSQTIAKMSYLEEKRLVHRDLAARNVLV 1071
QY 848 KSPNHVKITDFGLARLLDIDETEHADGGKVKPIKMALESILRRFTHQSDVMSYGVTVW 907
Db 1072 QTPSLVKITDFGLAKLLSSDSNEYKAAGKPKIKWALECIRNVRTSKSDVMAFGVTIW 1131
QY 908 ELMTFGAKYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCMIDSECRPRFREL 967
Db 1132 ELLTFGORPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCWHLDAAMRPTFKLT 1191
QY 968 SEFSRMARDPQRFVVIQNEGLG--PASPLDSTFFVRSLEDD---DMGLDVAEYLVPOQ 1022
Db 1192 TVFAEFADPGRYLAIPGDKFTRLPA-----YTSQDEKOLIRKLAPTTPDGEAIAKPD 1244
QY 1023 GFFCPDPAPGAGGMVHHRRSSSTRSGGDLTLGLLEPSEERAP-----RGLPAPSEGAG 1076
Db 1245 DYLQPKAAPGPS-----HRTDCT-----DEMFKLNRCKDPSNKNSTG 1283
QY 1077 SDVFDG----DLGMGAAGLQSLPHTDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEY 1133
Db 1284 DDERDSSAREVGVGNLR-----LDLPVDEDDYLMPTCQPGNN 1321
QY 1134 VNQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKVAFAGGAVENPEYL-- 1191
Db 1322 NNNNN-----NPNQNNAAVGAAGY-----DLIGVPVSDNPEYLLN 1360
QY 1192 --TPQGGAAPOPH-----PPPAFSP-AFDNLYYMD 1218
Db 1361 AQLTGVGESPIPTQITIGIPVNGGPGTMEVKVMPGSEPSTSSDHEYND 1408

RESULT 12
ERBB ALV
ID ERBB ALV
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
```

```
OS Avian leukosis virus.
OC Viruses; Retrovirdae; Retroviridae; Alpharetrovirus.
RN NCBI_TaxID=11864;
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RT Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; M10066; AAA48763.1; ALT_INIT.
CC PIR; A00643; TVCHLV.
CC PIR; B00643; TVFVLV.
CC HSSP; P11362; 1EGK.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; TyrKc_1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
CC Transfaser; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC KW Glycoprotein; Phosphorylation.
CC FT DOMAIN 132 399 PROTEIN KINASE.
CC NP BIND 138 146 ATP (BY SIMILARITY).
CC FT BINDING 165 165 ATP (BY SIMILARITY).
CC FT ACT SITE 257 257 BY SIMILARITY.
CC SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.1%; Score 1706.5; DB 1; Length 634;
Best Local Similarity 51.6%; Pred. No. 5.2e-83;
Matches 364; Conservative 73; Mismatches 142; Indels 127; Gaps 16;

QY 587 CAHYKDPFCVPCSPGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCPCAQ 646
Db 3 CAHIDGPHCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPNCCTRGCKPGLEGCP--- 58
QY 647 RASPLTSIQYTKANSKF-----IGITELKRRQOKIRKYVTRRLLOETELVEPLTPSGAM 700
Db 59 NGSKTPSIAGVGVGGLLCLLVVVGIGLYGLYLRHVRKRTLLRLLQERLVEPLTPSGEA 118
QY 701 PNOAMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIIPAIVKURENTSPKANKE 760
Db 119 PNOAHLRIKETEPKVKVLGSGAGFTVYKGLWIPDGENVKIIPAIVKURENTSPKANKE 178
QY 761 ILDEAYVMAGVSPYVRLGLICLTSTVQLVTQMPYGCGLDVRNRCGLCSODLLNWC 820
Db 179 ILDEAYVMASVDNPHVCRLLGICLTSTVQLITQMPYGCGLDVRNRCGLCSODLLNWC 238
QY 821 MQIAKMSYLELDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVP 880
```

```

Db 239 VOIAGMNYLEERRLVHRDLAARNVLKTPQHVKITDFGLAKLGADEKEYHAEGGKVP 298
Qy 881 KMALESILRRFTHQSDVWSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPP 940
Db 299 KMALESILHRIYTHQSDVWSYGVTVWELMTGSKPYDGPASEISSVLEKGERLPQ 358
Qy 941 CTIDVYMWKCMWIDSECRPFRELVSFSMARDPQRFVVIQ-NEDLGASPLDSTFY 999
Db 359 CTIDVYMWKCMWIDADSRPFRELIAEFSKWARDPPRYLVIOGDERMHLSPDTSKY 418
Qy 1000 RSLLEDDDDMDGLVDAAEYLVLPQGFCDPAPACAGMWHRRSSSTRSGGDLTLGLEP 1059
Db 419 RTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1060 SEEEAPRSL-----APSEGASDVDFDGLGMAAGLQSLTHDPSPLQRYSEDTPVL 1114
Db 450 -----SRTPLLSLSLTSATNSATNCID-----RNGQGHFVREDSFQRYSSDPTGNF 496
Qy 1115 PSET--DGVAFLTCSPQPEYVQDVRPQPPSPREGPLPAARPGATLERAKTILSPGKN 1172
Db 497 LEESIDDDGFL-----PAPEYVQ--LMPKKSTAM----- 524
Qy 1173 GVVKDVF-----AFGAVENPEYLTPOGGAAPQPHPPAFSPAFONLYYM 1217
Db 528 QIYNNISLTAISKLPMSDRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSPYW 579
Qy 1218 DQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGIDVP 1248
Db 580 IQSGNHQINLNDPQYQDPLNETKPNGLLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB AVIER
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1] SEQUENCE FROM N.A.
RP STRAIN=H;
RC MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Deubire B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1;
DR PIR; A00644; TVYUH.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76BECDD06745D609 CRC64;

Query Match 24.4%; Score 1660; DB 1; Length 604;
Best Local Similarity 51.4%; Pred. No. 1.4e-80;
Matches 354; Conservative 70; Mismatches 135; Indels 130; Gaps 15;

Qy 587 CAHYKDPFPCVARCPGKVPDLSPYMPKFPDEEGACQPCPNCINCHSDVLDKGCPEAQ 646
Db 3 CAHFIDGHCVCACAGVLGENDTL-VKAYADANAVCQLCHPNCCTGCGKPGLEGCP--- 58
Qy 647 RASPLTSIQYIKANSKF-----IGITELKRRQOKIRKYTMRLLOTELVEPLTPSGAM 700
Db 59 NGSKTPSTAAAGVVGGLCLLVVGLGIGLYLRRRHIVRKETLRLLQERELVEPLTPSGEA 118
Qy 701 PNAQMRILKETELRKVKVLGSGAGTGYKGIWPDGENVKIPVAIKVRENTSPKANKE 760
Db 119 PNAQHLRLKETEFKKVKVLGSGAGFTYKGLWIPEGEKVKIPVAIKELREATSPKANKE 178
Qy 761 ILDEAYVMAGVGSPPVSRLLGICLTSTVOLTPQMPYGLLDHVRNRCRLGSGDLLNWC 820
Db 179 ILDEAYVMASVDNPHVCRLLGICLTSTVOLITQMPYGLLDYIREHKONIGSQYLLNWC 238
Qy 821 MQIAGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 880
Db 239 VOIAGMNYLEERRLVHRDLAARNVLKTPQHVKITDFGLAKLGADEKEYHAEGGKVP 298
Qy 881 KMALESILRRFTHQSDVWSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPP 940
Db 299 KMALESILHRIYTHQSDVWSYGVTVWELMTGSKPYDGPASEISSVLEKGERLPQ 358
Qy 941 CTIDVYMWKCMWIDSECRPFRELVSFSMARDPQRFVVIQ-NEDLGASPLDSTFY 999
Db 359 CTIDVYMWKCMWIDADSRPFRELIAEFSKWARDPPRYLVIOGDERMHLSPDTSKY 418
Qy 1000 RSLLEDDDDMDGLVDAAEYLVLPQGFCDPAPACAGMWHRRSSSTRSGGDLTLGLEP 1059
Db 419 RTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1060 SEEEAPRSL-----APSEGASDVDFDGLGMAAGLQSLTHDPSPLQRYSEDTPVL 1114
Db 450 -----SRTPLLSLSLTSATNSATNCID-----RNGQGHFVREDSFQRYSSDPTGNF 496
Qy 1115 PSET--DGVAFLTCSPQPEYVQDVRPQPPSPREGPLPAARPGATLERAKTILSPGKN 1172
Db 497 LEESIDDDGFL-----PAPEYVQ--LMPKKSTAM----- 524
Qy 1173 GVVKDVF-----AFGAVENPEYLTPOGGAAPQPHPPAFSPAFON 1213

```

```

Db      525 -VQNIQYINFIISLTALSKULPMDSRQVNSHSTAVDNPEYL-----NTNQSPKLVTFES 578
Qy      1214 LYYWDQDPPERGAPPSTFKGTPTAENPEY 1242
Db      576 SPYMIQSNHQ-----INLNDPDY 594

RESULT 14
ID      ERBB_AVIEU          STANDARD;          PRT;          540 AA.
AC      P11273;
DT      01-JUL-1989 (Rel. 11, Created)
DT      01-JUL-1989 (Rel. 11, Last sequence update)
DE      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN      V-ERBB.
OS      Avian erythroblastosis virus (strain ts167).
OC      Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX      NCBI_TaxID=103898;
RN      [1]
RX      SEQUENCE FROM N.A.
RX      MEDLINE=87064458; PubMed=2878364;
RA      Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RA      "A single amino acid substitution in v-erbB confers a thermolabile
RT      phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT      cells.";
RT      Mol. Cell. Biol. 6:1751-1759(1986).
CC      -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
EMBL; M13179; AAA42401.1; -.
DR      PIR; A25231; TVFVEB.
DR      HSP; P11362; 15GK.
DR      InterPro; IPR000719; Euk_pkinase.
DR      InterPro; IPR001245; Tyr_pkinase.
DR      Pfam; PF00069; pkinase; 1.
DR      ProDom; PD000001; Euk_pkinase; 1.
DR      SMART; SM00219; TyrKc; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW      Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW      Glycoprotein; Phosphorylation.
FT      DOMAIN 132 399 PROTEIN KINASE.
FT      NP_BIND 138 146 ATP (BY SIMILARITY).
FT      BINDING 165 165 ATP (BY SIMILARITY).
FT      ACT_SITE 257 257 BY SIMILARITY.
FT      VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ      SEQUENCE 540 AA; 60412 MW; 5B53297AA068B65D CRC64;

Query Match 23.4%; Score 1587; DB 1; Length 540;
Best Local Similarity 54.2%; Pred. No. 8.5e-77;
Matches 335; Conservative 63; Mismatches 126; Indels 94; Gaps 13;

Qy      587 CAHYKDPFFVCARCPGVKPDLSVMPYTKFPDEGACQPCPINTCHSVLDLDDKGCRAEQ 646
Db      3 CAHFIDGPHCVKACPAVLGENDTL-VNKYADANAVCOLCHPNCTRGCKGPGLEGCP--- 58
Qy      647 RASPLTSIQIYKANSKF-----IGITELKRQOKIRKYTMRRLLQETTELVEPLTPSGAM 700
Db      59 NGSKTPIAAGVGVGGLLCLVWVGLIGLYLRRRHIVRKRTLRLQLQRELVEPLTPSGEA 118
Qy      701 PQNAQRRLKETELRKVKVLGSAFGTVYGIWIPDGENVKIPVAIKVLRNTPSKANKE 760

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-369-383-12

Perfect score: 6815

Sequence: 1 MELAALCRWGLLLALLPRGA.....TFKGTPTAENPEYLGLDVVP 1255

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriaph.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6152	90.3	1259	6	O18735 canis famil
2	3096	45.4	1209	11	Q9GX70 rattus norv
3	3067	45.0	1210	11	Q9EP98 mus musculu
4	2692	39.5	1165	13	Q9YH40 xiphophorus
5	2645.5	38.8	1137	13	Q9W6F6 gallus gall
6	2264	33.2	1328	13	P79754 fugu rubrip
7	2010.5	25.5	1433	5	Q9BIH9 anopheles g
8	1871	27.5	419	4	Q9UK79 homo sapien
9	1739	25.5	367	11	Q8R2X1 mus musculu
10	1720	25.2	729	15	Q86712 avian rous-
11	1718	25.2	567	15	Q86714 avian rous-
12	1697.5	24.9	412	4	Q8WV0 homo sapien
13	1662.5	24.4	962	15	Q64895 avian eryth
14	1645	24.1	545	15	Q85468 avian eryth
15	1458.5	21.4	655	11	Q9WVF5 mus musculu
16	1442.5	21.2	643	11	Q9ERV6 mus musculu

Q9YIX8 ephydatia f
Q23821 caenorhabdi
Q26566 schistosoma
Q90836 gallus gall
Q9ese0 rattus norv
Q14256 homo sapien
Q9psn2 gallus gall
Q923v5 rattus norv
P11776 xiphophorus
Q8szw1 drosophila
Q99162 xiphophorus
Q9pvz4 xenopus lae
Q9bud7 homo sapien
Q9njv5 biophalari
Q9bg66 oryctolagus
Q93457 scophthalmu
Q8uw85 paralichthy
Q8uw86 paralichthy
Q9u5a8 bombyx mori
Q73798 xenopus lae
Q8uw84 paralichthy
Q8uw83 paralichthy
Q9qv4 rattus sp.
Q9y9h8 scophthalmu
Q9vds4 drosophila
Q96135 homo sapien
Q07912 homo sapien
Q54967 mus musculu
Q9umq4 homo sapien

ALIGNMENTS

RESULT 1

O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Erbb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]_TaxID=9615;
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbb-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1;
DR HSSP; P11382; 1FGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match		90.3%;	Score 6152;	DB 6;	Length 1259;
Best Local Similarity		90.0%;	Pred. No. 0;		
Matches 1134;		Conservative 45;	Mismatches 75;	Indels 6;	Gaps 2;
QY	1	MELAAALCRGGLLALLPPGAAS	TOVCTGTDMLKRLPAS	PETHLDMLRHLYQG	COVQGNL 60
DB	1	MELAAWCRGGLLALLPPGAAGT	TOVCTGTDMLKRLPAS	PETHLDMLRHLYQG	COVQGNL 60
QY	61	ELTYLTNLSASLFLODIOE	VOGYVLIAHNOVQVPLQRL	IRVRGTQFEDNYALAVLDNG	120
DB	61	ELTYLTNLSASLFLODIOE	VOGYVLIAHNSQVRQIPLQRL	IRVRGTQFEDNYALAVLDNG	120
QY	121	DELANTPTVTGASPGGLRE	LQSLTEILKGGVLIQBNPOL	CVQDTILKXIDIFHKKNOLA	180
DB	121	DELGGIPAPGAAGGLRE	LQSLTEILKGGVLIQBNPOL	CVQDTILKXIDIFHKKNOLA	180
QY	181	LTLIDNRSRACHPCSPMK	CGSESSDCSLTRTV	CAGGCARCKGPLPTDC	CHEQC 240
DB	181	LTLIDNRSRACHPCSPACK	DAHCWAGSSDCSLTRTV	CAGGCARCKGQPPPTDC	CHEQC 240
QY	241	AAGCTGPKHSDCLACLFH	NHSGICELHCPALVTYNTD	TFESMPNPEGRYTFGASC	VTACP 300
DB	241	AAGCTGPKHSDCLACLFH	NHSGICELHCPALVTYNTD	TFESMPNPEGRYTFGASC	VTSCP 300
QY	301	YNYLSTDVSGSCTLVCP	LHNQEVTAEDGTORCEK	SPCARVCYGLGMQYIKANS	KFIGIT 360
DB	301	YNYLSTDVSGSCTLVCP	LHNQEVTAEDGTORCEK	SPCARVCYGLGMHLEVR	AVTSAN 360
QY	361	ELEFACKKIFGSLAF	LPESFDGDPASNTAPOYI	KANSKFIGITELTVYLI	ISAWPDSL 420
DB	361	IOEFACKKIFGSLAF	LPESFDGDPASNTAPLO	EQURVFEALIEITGVYLI	ISAWPDSL 420
QY	421	DLVSFQNLQVIRGRIL	NHNGAYSITLQGLG	ISWGLSLRELGSGLAL	IIHNTHLCFVHTV 480
DB	421	NLSVFQNLVIRGRVLD	GAYSITLQGLG	ISWGLSLRELGSGLAL	IHRNARLCFVHTV 480
QY	481	PHDQFRNPHQALHTAN	PEDECVGEGLAGHOL	CARGHCWGPGTQVCNCS	OFLRGQEC 540
DB	481	PHDQFRNPHQALHSAN	PEEBCVGEGLACYP	CAHGHCMGPGPTQVCNCS	OFLRGQEC 539
QY	541	VVECVRLQGLPREYNAR	CLPCHPECPQNGSVT	CFGPEADOCVACAHYK	DPFPCVARC 600
DB	540	VVECVRLQGLPREYV	KDRVCLPCHSECPQNGSVT	CFGSEADOCVACAHYK	DPFPCVARC 599
QY	601	PSGVKPDLSYMPITW	KFPDEGACQPCPINCTH	SCVDLDDKCPAEQASPLTS	IVSAVVG 660
DB	600	PSGVKPDLSYMPITW	KFADEEGTCQPCPINCTH	SCADLDEKCPAEQASPLTS	IIAAVVG 659
QY	661	ILLVVLGVVFGILIK	ERQOKIRKYTMRLLQ	ETELVEPLTPSGAMPNQA	OMRILKETEL 720
DB	660	ILLVVLGVVFGILIK	RRRQKIRKYTMRLLQ	ETELVEPLTPSGAMPNQA	OMRILKETEL 719
QY	721	RKVVLGSGAFGVYK	GIWIPDENYKIPVAIK	VLRENTSPKANKEILDE	AYVMAGVGP 780
DB	720	RKVVLGSGAFGVYK	GIWIPDENYKIPVAIK	VLRENTSPKANKEILDE	AYVMAGVGP 779
QY	781	YVSRLLGICLTSTV	QLVTLMPYGCILLD	HVRENRGLSGQDLLN	WCMIKAGMSYLEDVR 840
DB	780	YVSRLLGICLTSTV	QLVTLMPYGCILLD	HVRENRGLSGQDLLN	WCMIKAGMSYLEDVR 839
QY	841	LVRDLAARNLVKSP	NHVKITDFGLARLLD	IDETEHADGGKVP	IKWMALESILRRFT 900
DB	840	LVRDLAARNLVKSP	NHVKITDFGLARLLD	IDETEHADGGKVP	IKWMALESILRRFT 899
QY	901	HQSDVMSYGVTV	WELMTFCAKPYDGI	PAIREIPDLLEKGERL	POPPICTDVVMYKCM 960
DB	900	HQSDVMSYGVTV	WELMTFCAKPYDGI	PAIREIPDLLEKGERL	POPPICTDVVMYKCM 959
QY	961	IDSECRPRELVSE	FSRWARDQRFVITQ	NEDLGASPLDSTFY	RSLLDEDDMGDLVDA 1020
DB	960	IDSECRPRELVSE	FSRWARDQRFVITQ	NEDLGASPLDSTFY	RSLLDEDDMGDLVDA 1019
QY	1021	ESYLVPOQGF	FFCPDPAPAGGVMV	HRHSSTSRSGGDL	ITGLPESEEEAPSLAPSEG 1080

DB	1020	EBYLVPOQGF	FCPEPTGAGGTA	HRHRSSSTRNGGELT	ITGLPESEEPKSLAPSEG	1079
QY	1081	AGSDVFDGDLG	MGAAKGLQSLP	THDPSPLQRYSED	PTVPLPSETDCGYVAPLTCSPQPEYV	1140
DB	1080	AGSDVFDGDLG	MGAAKGLQSLP	SDQPSPLQRYSED	PTVPLPETDGRKVAPLTCSPQPEYV	1139
QY	1141	NQPDVVRPQPPS	PREGGLPAARPA	GATLER-----AKT	LSFGKNGVVKDVFAGGAVENPE	1195
DB	1140	NOPEVMPQPLA	EGPLPPSRPAGAT	LERPKT	LSPKLTSPGKNGVVKDVFAGGAVENPE	1199
QY	1196	YLTPOGGAAPQ	PHPPAFSPAF	DNLYYWDQDP	PERGAPSTFKGTPTAENPEYGLDVPV	1255
DB	1200	YLAPRGAAPQ	PHPPAFSPAF	DNLYYWDQDP	SERGSPPSTFEGTPTAENPEYGLDVPV	1259
RESULT 2						
QY	Q9QX70	PRELIMINARY; PRT; 1209 AA.				
ID	Q9QX70;					
AC	Q9QX70;					
DT	01-MAY-2000	(TremBLrel. 13, Created)				
DT	01-MAY-2000	(TremBLrel. 13, Last sequence update)				
DT	01-JUN-2002	(TremBLrel. 21, Last annotation update)				
DE	Epidermal growth factor receptor.					
GN	EGFR.					
OS	Rattus norvegicus (Rat).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
OX	NCBI_TaxID=10116;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=FISHER; TISSUE=LIVER;					
RX	MEDLINE=90258888; PubMed=2342466;					
RA	Petch L.A., Harris J., Raymond V.W., Blaeband A.J., Lee D.C.,					
RA	Earp H.S.;					
RT	"A truncated, secreted form of the epidermal growth factor receptor is					
RT	encoded by an alternatively spliced transcript in normal rat tissue.";					
RL	Mol. Cell. Biol. 10:2973-2982(1990).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=FISHER; TISSUE=LIVER;					
RA	Petch L.A.;					
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.					
RN	[3]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=FISHER; TISSUE=LIVER;					
RA	Guttridge K., Dawson T.L., Earp H.S.;					
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; M37394; AAF14008.1; -.					
DR	HSSP; P11362; 1FGK.					
DR	InterPro; IPR000494; EGFR_L domain.					
DR	InterPro; IPR000719; Euk_pkinase.					
DR	InterPro; IPR002174; Furin-like.					
DR	InterPro; IPR001245; Tyr_pkinase.					
DR	Pfam; PF00757; Furin-like; 1.					
DR	Pfam; PF00069; pkinase; 1.					
DR	Pfam; PF01030; Recep_L domain; 2.					
DR	PRINTS; PR00109; TYRKINASE.					
DR	ProDom; PD000001; Euk_pkinase; 1.					
DR	SMART; SM00261; FU; 3.					
DR	SMART; SM00219; TyrK; 1.					
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.					
DR	PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.					
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.					
KW	ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.					
SQ	SEQUENCE 1209 AA; 134891 MW; 96FEE7F6CC1B7773 CRC64;					
Query Match 45.4%; Score 3096; DB 11; Length 1209;						
Best Local Similarity 49.9%; Pred. No. 1.2e-224;						
Matches 638; Conservative 162; Mismatches 362; Indels 116; Gaps 28;						
QY	3	LAALCRWGLLLALL	PPGA-ASTQVCTG	TDMLKRLPAS	PETHLDMLRHLYQG	VOVQGNUE 61

Db 15 LAALCAAG-----GALEBKVKCGQTSNRLTQLTGFEDHFLSLQRMFNFNCEVVLGNLE 66
Qy 62 LTYLPTNASLFLQDIOEQVQYVLIANVQVQVPLQRLRIVRGQTQLFEDNTALAVLQNGD 121
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 124
Qy 67 ITVQRNYDLSFLKIQEVAGYVLIATNTVERIPLNLIQIRGNALYENTYVALAVLSN-- 124
Qy 122 PLNNTTPVTGASPGGLREQLRSLEILKGGVLIQRPOLCVQDTILWKDIFHKNLQAL 181
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 175
Qy 182 TLIDTNRS-RACHPCSPMKSGRCWGESSEDCQSLTRTVACAGCA-RCKGGLPLTDCCHQ 239
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 235
Qy 240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCTAC 299
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 295
Qy 300 PNYLSTDVSGSCTLVCPHNOVEATDGTORCEKSPCARVCYGLGMOYIKANSKFIGI 359
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 353
Qy 360 TELE-FAGCKKIFGSLAFLPESFDGDPASNTAPQYIKANSKFIGITELTYLYISAWPDS 418
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 413
Qy 419 LPDLSVFQNLQVIRGRILHNGAYSLSLTQGLGISWGLRSRLBELSGSLALIHNTLFCVH 478
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 473
Qy 479 TVPMDQLFRPHQALLHTANRPEDECVEGLACHOLCARGHCWGPPTQCVNCSQFLRGQ 538
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 533
Qy 539 ECVEECRVLOGLPREYVYARHCLPCHPECPQNGSVTCFGEADQCACAHYKPPFCVA 598
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 593
Qy 599 RCGSVKVPDLSYMPIWKFPEEGACQCPINCTHSCVDLDDKGPASQORASP-LTSTVSA 657
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 650
Qy 658 VGGILLVVLGWFGI-LIKRROQKIRKYMRLLELVEPLTPSGAMPNQOQMRILK 716
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 709
Qy 717 ETELKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKIILDEAYVMAG 776
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 769
Qy 777 VGSPIYSRLIGICLTSTVOLVTQMLPYGCLLDHVRENRGLSGODLLNWCQIAKMSYL 836
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 829
Qy 837 EDVRLVHRDLAARNVLKSPNHVKITDIFGLARLLIDITETVHADGGKVPKKNMALESILR 896
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 889
Qy 897 RRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYIMV 956
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 949
Qy 957 KCMWIDSECPRELVSFSRWARDQRFVWIQ-NEDLGPASPLDSTFVRSLLDDDMG 1015
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1009
Qy 1016 DLVDAEYLVPOGFFCPDPAPGAGVHHRHSSSTRSGGDLTLGLEPSEEAAPSPL 1075
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1035
Qy 1076 APSEGAGSDVFDGDLGMAAKQLSLPHTDPSPLQRYSEDPTVPLPSET--DGVAPLTC 1133
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1086
Qy 1036 LSSLANSN----SSTVACINRNGSCRVKEDAFIQRYSSTPTSVLTEDNTDITFL-----

Qy 1134 SPQPEYVNPQVRRPQSPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKVDFAFGAVEN 1193
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1134
Qy 1194 PEYL-TPOGGAAPQHPHPPAFSPADNLVYNDQ-----DP-----PERGAPPST 1236
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1185
Qy 1237 FKGTPTAENPEYLGLDVP 1254
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1202
Qy 1254 FKGTPTAENPEYLGLDVP 1254
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1202

RESULT 3
Q9EP98
ID Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RC SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Fearall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J;
RC Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RC Schehl C., Fearall R.S., Green P.J., Yee D., Lampland A.L.,
RC Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RC Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAG28045.1;
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG24386.1;
DR HSP; P11362; IFGK.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR001719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00130; CYTOCHROME_C; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

Db 3 LCFADTHWQDIVRNPWASNTFLVPTNGSSGCRCHKSCGTG-RCWGPTENHCQTLLTKVC 61
QY 221 AGGC-ARCKGPLPTDCCHCCAACTGPKSDCLACLFHNSHIGICELHCPALVTYNTDTF 279
Db 62 AEQCDGRCYGPVSDCHRECAAGCGPKDTCFACMNFDSGACVTCQPTOTFYVNPPTF 121
QY 280 ESMENPGRVTFGASCVTACPYNYLSDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCA 339
Db 122 QLEHNNHAKYTYGAFCKVCKPHNFV-VDSSSCVRACPSRKEV-EENGKWKCKPCTDIP 179
QY 340 RVCVGLGMQYIKANSKFIGITELE-FAGCKIFGSLAPLPSPFGDDPASNTAPQYIKANS 398
Db 180 KACDGICTGSL-VSAQTVDSNDKFNCTKINGNLIFLVTGIGHDPVHTIAANPEKLN 238
QY 399 KFIGITELTGVLYLSAMPDLSLDFQNLQVIRGRILHNCAYSILTLQGLISWGLRSL 458
Db 239 IFQTVREITGVNLQSPENMTDFRFSNLVTIGRALYSLGLSLLILKQOQITSLQFQSL 298
QY 459 RELGSLALIHNTHLCFVHTVPWDQLFRNPHQALLHTANRDECEVGEGLACHQLCARG 518
Db 299 KQISAGNIYITDNSLNCYHTVNTSLSFTSPQKTVIHRNKKAEKNTADGMVNCNELSSD 358
QY 519 HCWGPPTQCVNCQFIRGQSCVEECRVQLGRLPREYVNAHCLPCHPEQCP-ONGSVTFC 577
Db 359 GCWGPDPQCLSKCFIRGRCTIESCNLYDGEFREFANGSVCMCEDPQCEKMDNMITY 418
QY 578 GPEADQCVACAHYKDPFCVACRCSGVKPDLSYMPIMKFPDEEACQPCPINCHTSCVDL 637
Db 419 GPGPDHCTKCFHFGDGNCEKCPDGLQGANF--IFKYADEDECHPCPNCTQGGCRGP 476
QY 638 DDGCG-----PAERASPLTSIVSAVV-GILLVVLGVVFGILIKRROKIRKY 685
Db 477 ASHDCIYVPTWQSTLPOHAR-TEL--IAAGVIGLFIIVIMGLTFAVYVRKSIK-KKR 532
QY 686 TMRLLQETELVEPLTSGAMPNOAQMRILKETELRKVKULGSCAFGTGVKGIWPDGEN 745
Db 533 ALRRL-ETELVEPLTSGTAPNOAQLRIKELTELKRVKVLGSGAFGTGVKGIWVPEGT 591
QY 746 VKIPVAIKVLRNTPRANKELDEAVYMGVGPVYRLLIGLICLTSTVOLVTOLMPYGC 805
Db 592 VKIPVAIKILNETGPKANVEFMEALIMASMDHPHVLRLGVCLSPITQLVLTOLMPHGC 651
QY 806 LLDHVRNRRGLGSDLLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFG 865
Db 652 LLDVYHEKDNIGSOLLNMCVQIAKGMVLEERLVRDLAARNVLKSPNHVKITDFG 711
QY 866 LARLLDIDETEHADGKVPKIMWALESIILRRRTHOSDVWSYGVTVWELMTFGAKPYDG 925
Db 712 LARLLEGEKEYNADGGKMPKIMWALECIHYRKFTHOSDVWSYGVTVWELMTFGKPYDG 771
QY 926 IPAREIPDLLEKGERLPPOPCTIDVYIMVWKMWIDSECRPRFRELVSFESRMARDPQR 985
Db 772 IPTREIPDLLEKGERLPPOPCTIDVYIMVWKMWIDSECRPRFRELVSFESRMARDPQR 831
QY 986 FVVIQNEED-LGPAASPLDSTFYRSLLEDMDGLVDAEYLVPOQGFCCPDPAAGAGGVH 1044
Db 832 YLVITQGDRLKPLSPNDKSFQNLDEEDLEMDMAEYLVLP-QAFNIPPIYTSRTRID 890
QY 1045 HRRSSSTRSGGDLTLGLSPSEEAAPRS--PLAP-SEGAGSVDFDGLGMAAGLQSL 1101
Db 891 SNRNQFVYRGGVAAEGV-PMPYRAGPCTIPEAPVAGATAEITFEDTCNGTLRKQVAT 949
QY 1102 PTHDPSPLQRYSEDPTVPLPS-----ETDGYVAPLTCSPQEPYVNPQDVRPQPPSPRE 1154
Db 950 LAKEDSSTORYSADPTVFIPEVIRGELDEGYTPMRDKPTDYLNPEENPVPVSRKN 1009
QY 1155 GPLPAA-RPAGATLERAKTISPGKGVVQKVF-----AFGAVENPEYLTPOGGAAP 1205
Db 1010 GDQAVDNPYHN-----APNQPKAEDEYVNEPLYLNTFANTLENAEYL----- 1054
QY 1206 QPHPPPPAFSAFONLYWQDDPPERGA--PPSTFKGTPT-----AENPEYL 1249

Db 1055 KNNLPEKAKKAFNDPDMYNNHSLPPRSTLQHPDYLQLEYSTKYFKQNGRIRPIVAENPEYL 1114
RESULT 6
P79754
ID P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ErDB3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gellner K., Brenner S;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
rubripes.";
RL Genome Res. 9:251-258(1999).
DR HSSP; AF056116; AAC34391.1; --
DR HSSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 1328 AA; 148613 MW; A33039258B647E9 CRC64;
Query Match 33.2%; Score 2264; DB 13; Length 1328;
Best Local Similarity 40.0%; Pred. No. 9.3e-162;
Matches 515; Conservative 155; Mismatches 425; Indels 192; Gaps 32;
QY 9 WGLLLALLPP--GAASQTQ-----VCTGDMKLRLPASPETHLDMLRLHYLQGVQVQVGNL 62
Db 4 WRLILMCVASRLRAASSQTQEA VCPGTQNGLSSTGSQENQYNLKNDRYKGCETIIMGNLEI 63
QY 63 TYLPTNASLSFLQDIOEVQGVLIHAHNOVQPLQRLRIVRGTLFEDNVALVLDNGDP 122
Db 64 TQESNWDFSLKTIREVGTGVLAMNHFOEIPUGQURVIRGNSLYERRRFPALSVFLN--- 120
QY 123 LNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILMKDIFHKNQLALT 182
Db 121 ----YPKDG--PSGLNGLMLNLTLLDGGVLIINNKYLRYPVWYRDII--RNDAPIE 173
QY 183 LIDTNSRACHPCSPMKGRSCWGESSEDCQSLTRTVACGC-ARCKGPLPTDCHEOCA 241
Db 174 IQFNGERGVCH---KSC-GNYCWPGRKQCCQIILTKVCAPOCNDRCFCGTSRDCCHIECA 229
QY 242 AGCTGPKHSDCLACLFHNSHIGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPY 301
Db 230 AGCKGPLDTQCFACRLFNDGACVPQCPQTLIYNKQTFQMETNPNAKTYGSIQVSCQPT 289
QY 302 NYLSTDVGSCTLVCPHLNQEVTAEDGTQR-CEKSKPCARVCYGLGMQYIKANSKFIGIT 360
Db 290 HFV-VDGSSCVSVCPDPKMEV--ERGSQRCCLSGCLCPKVCETGAE----QRQTVDS 342
QY 361 ELE-FAGCKIIFGSLAPLPSPFGDDPASNTAPQYIKANSKFIGITELTGVLYISAWPDSL 419
Db 343 NIDSFINCTKIQSLHFLVLTGILGDDPKVNPPLDAKLEVFRTVREITDILNIQSWPKEL 402

```
QY 420 PDLVSFQNLQVIRGRILHNGAYSLSLTQGLGSLWGLSLRSLRELGLALIHHTLCPVHT 479
DB 403 NDLVSFSLTTIQGRSLFRFSLVMVRIPLTLSLGLSLRSLREISDGSVYISQNAHLCLVHT 462
QY 480 VPMQDLFRNPH-QALLHTANRDECEVGEGLACHQLCARGHCGPGPTQCVCNCSQFLRGQ 538
DB 463 VNMVQLFRGRSRVANSNSRNPMAECVADGRVCDPLCSGCGGPGPDQCLSRNSRHG 522
QY 539 ECVECRVLOGLPREYVNAH-CLPCHPECPQNGSVTCFPGADOCVACHAKYKDPDFCV 597
DB 523 TCVAGCHFNSGIPREFAGLNGVCVACHPECKPQTGRASCTGPGADECMACTKFRDGYCM 582
QY 598 ARCPGSGVKPDLSPYMTWKPDEGACOPCPINCTHSCVDLDDGCPABQASPLTSVISA 657
DB 583 SSCPAGVN-DGEKGLIFKFPNREHCEPCQNCTQCGSGPLNDC---LEAARLTSSGQ 638
QY 658 VVGILLVWLVGVF-----GILIKRQOKIRKYTMRRLLQBELTELVEPLTPSGAMPNOA 710
DB 639 ITGIALGVPAGLIFCLVFLGLMYHRLAIRKRAMRYLESGESFEPLGP-GEKTKV 697
QY 711 QMRILKETELRKVKVLGSGAGFTVYKGIWTPDGENKIPVAIKVLRNTSPKANKELDE 770
DB 698 HARILKPSDLRKIKPLGSGVFGTVSGFWIPEGETVKIPVAIKTIQDSSGRQTFTEIDH 757
QY 771 AYVMAGVGSPPYVSRLLIGTCLTSTVOLVTQMLMPYGCILLDHVRENRLGSGQDILLNWCQIA 830
DB 758 LLSMGLSDHPYIVRLGICPGTCLQVLTQSSHSLLEHRIROHKTSLDPORLNWCQIA 817
QY 831 KGSYLEDVRLVHRLDAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVKVTKMMA 890
DB 818 KGMVYLEHRVHVHKNLAARNILLKNDYQVQISDYGVADLLYPDKKVYVSETKTPKMA 877
QY 891 LESILRRRTHQSDVMSGYVVMELMTFGAKPYDGIPIAREIPOLLEKGERLPQPICTID 950
DB 878 LESILFRYTHQSDVMSGYVVMWMSFGAEPYASVQPEVPSVLEKGERLSQPAICTID 937
QY 951 VYMVWKCMWIDSECPREFLYSEFSRMAARDQRFVVIQNEDLGPASPLDSTFYKSLLE 1010
DB 938 VYMVWKCMWIDENIRPTFKELASDFTRWARDPPRYLVIRMEG-----E 981
QY 1011 DDDMGDLVDAEEYLVQGGFFCPDPAPAGGVMVHRRSSSTRSGGDLTLGLEPSEEA 1070
DB 982 DSGMGFL-----RGRSER---GLLEADLEDEEE- 1008
QY 1071 PRSPLAPSEGASDVDFDGLGMG---AAKGLQLTHDPSPLO-----RYSDDPT 1117
DB 1009 -----GLDRFATPSLOPSPSWSTSPSINSVMVMTQLRYD----- 1044
QY 1118 VPLPSETDGVAPLTCSPQ- EYVQ-----PDVRPOPPSPREGPL--PAA 1160
DB 1045 --FAVSQGGHIGYLPMSPPVDITRQLWYQSRSLSSVRLPDRSAFRSSREAELEDGA 1102
QY 1161 RPAGATLERAKTILSPCKNGVVKDVFAGGAVENPEYLTPOGGAAPQHPPPAPFAPDNL 1220
DB 1103 QCAGIFRVR-----FGSERGN-----PQGG----- 1122
QY 1221 YVWDQPPPERGAPPSTFKGTPTAENPE 1247
DB 1123 ---QQRKLSTASSPSSFKTWADEDEE 1146
```

RESULT 7

```
Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
```

```
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SUA;
RA Lyceet G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P11362; IFGK
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR P-coDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; Tyfkc; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER 1
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;
```

Query Match 29.5%; Score 2010.5; DB 5; Length 1433;
Best Local Similarity 32.8%; Pred. No. 1.4e-142;
Matches 471; Conservative 195; Mismatches 396; Indels 385; Gaps 38;

```
QY 26 CTGDTMKRLPASPETHDMLRHLYQCQVQGNLELTYPNTNASLFLQDIOBQGVYL 85
DB 1 CIGTNGRMSVPANREYHYKNLRDRYNTCTYVDGNLEITWIONITDLNLFHIREVTGYL 60
QY 86 IAHNVQVPLQRLRIVRGTLF-----EDNYALAVLDNGDPLNNTTPTVTGASPGGLREL 140
DB 61 ISLYDLQVILPLQIIRGRTTFKLNKEAYGLFV-----SFSHMTL 104
QY 141 QLSRLTEILKGGVLIQNPOLCYQDTILWKDI-FHKNQALALTLDITNRSRACHPCSPMC 199
DB 105 ELPALRDIILGSGVGFNNYNLCHKMSINWEEIILAPQTSMOYTFNFSSPERVCPCHPSC 164
QY 200 KGRSCWGESSEDCQSILTRTVACGGCA--RCKGPLPTDCHEQCAAGCTGPKHSDCLACLH 257
DB 165 EVG-CHWEGAHNCQRFSLKNCSPQCSQGRCFGPKPRECHLFCAGGCTGTQSDCLACKN 223
QY 258 FNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNYVLTSDVGSCVLCPL 317
DB 224 FYDDGVCKQECPPMQIYNPTNFWEPNPGKAYGATCVRKCP-EHLLKONGACVRKCPK 282
QY 318 HNOEVTAEQGTQCEKCKSPCARVCYGLGMQYIKANSKFIGITELEFAGCKIFGSLAFL 377
DB 283 GRMPQNSE-----CVPECKGVCPKTCPEGGI-----VHSDNIG-----NYKDCITLEGSLLEIL 329
QY 378 PESFDGDDPASNT-----APOVIKAN-----SKFIGITELTGYLYISAWPDSLPDLVSFQNLQ 429
DB 330 DQSGDFGQQVYTNFSGPRYIKIDPRLEVFSTVKEITGFINQAHHPNTTLYNFANLE 389
QY 430 VIRGRILHNGAY-SLTQGLGISWLGLRSLRELGLALIHHTLHLCFVHTVPMQDLFRN 488
DB 390 VVGGRLKENLPASVYIVKTSLSLELKRWNSSIVILENSDLCFDEIDWSEIKKS 449
QY 489 PHQALLHTANRDECEVGEGLACHQLCARGHCGWPGPTQCVCNCSQFLRGQECVEECRVLQ 548
DB 450 SDHEVMVQKNRNATECHEEGMECSQSKAGCWGKGPEQCLECKKNYKYGKCLDSCK--- 506
```



```
Query Match      25.5%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 5.9e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

Qy 889 MALESILRRRTHOSDVMSYGVTVWELMTFCAGPYDGIIPAREIPDLLEKGERLPQPPICT 948
Db 1 MALESILRRRTHOSDVMSYGVTVWELMTFCAGPYDGIIPAREIPDLLEKGERLPQPPICT 60

Qy 949 IDVYIMVIMVCMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSL 1008
Db 61 IDVYIMVIMVCMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSL 120

Qy 1009 LEDDMGDLVDAEYLVPQGFCDPAPGAGGMVHRHRSSTSRSGGDLTLGLEPSEE 1068
Db 121 LEDDMGDLVDAEYLVPQGFCDPAPGAGGMVHRHRSSTSRSGGDLTLGLEPSEE 180

Qy 1069 EAPRSLAPSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYV 1128
Db 181 EAPRSLAPSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYV 240

Qy 1129 APLTCSQPEYVNOQVVRPQPPSPREGPLPAARPAGATLERAKTLSPGKGVVWVDVFAFG 1188
Db 241 APLTCSQPEYVNOQVVRPQPPSPREGPLPAARPAGATLERAKTLSPGKGVVWVDVFAFG 300

Qy 1189 GAVENPEYLTPOCGAAQPHPPRAFSFADNLYYWDQDPPERGAPPSTFKGTPTAENPEY 1248
Db 301 GAVENPEYLTPOCGAAQPHPPRAFSFADNLYYWDQDPPERGAPPSTFKGTPTAENPEY 360

Qy 1249 LGLDVPV 1255
Db 361 LGLDVPV 367

RESULT 10
Q86712 ID Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE POLYPROTEIN.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSP; P03322; IAGS.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro.M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFE1D63 CRC64;

Query Match      25.2%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 4.5e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;
```

```
Qy 569 PONGSVTCFGEADOCVACAHYKDPFCVACRPSGVKPDLSYMPFWKPPDEGACQPCPI 628
Db 141 PEETATPKTGP--DHCMAKAFIDGPHCVACRCPAGVLGENDTL-VMKYADANAVCOLCHP 197

Qy 629 NCHTSCVDLDDKGCFAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKVTM 687
Db 198 NCTRGCKGPGLEGCP---NGSKTFSIAAGVVGGLCLVVGGLGILGLYLRRL-HIVRKRRL 253

Qy 688 RRLLOETELVEPLTPSGAMPNQAMRILKETELRKVKVLGSGAGFTVYKGIWIPDGVNK 747
Db 254 RRLLOERELVEPLTPSGEAPNQAHILKETEFKVKVVLGSGAGFTVYKGIWIPGEKVK 313

Qy 748 IPVAIKVLRENTSIPKANKEILDEAYVMAGVSPVYSRLGLICTSTVOLVTQLMPYGCGL 807
Db 314 IPVAIKELRENTSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQLMPYGCGL 373

Qy 808 DHVRENFRGLSGODLNNCMQIAKMSYLEDVRLVHRDLAARNVLVKSNNHVKITDFGLA 867
Db 374 DYIREHKNIDNGSYLLNNCVOIAKGMNLYBEERLVHRDLAARNVLVKTPOHVKITDFGLA 433

Qy 868 RLLDIDETEHADGCKVPIKMMALESILRRRFTHQSDVMSYGVTVWELMTFCAGPYDGI 927
Db 434 KLLGADKEKHAEGCKVPIKMMALESILHRIYHQSDVMSYGVTVWELMTFCSPYDGI 493

Qy 928 AREIPDLLEKGERLPQPICTIDVYIMVIMVCMIDSECRPRFRELVSFSESRMARDPQRF 987
Db 494 ASEISSVLEKGERLPQPICTIDVYIMVIMVCMIDADSRPKFRELIAEFSSKWARDPPRYL 553

Qy 988 VIQ-NEDLGPASPLDSTFYRSLLEDMDGDLVDAEYLVPQGFCDPAPGAGGMVHR 1046
Db 554 VIQGERMHLPSPDTSKFYRLMEEDMEDIVDAEYLVPHQGF----- 598

Qy 1047 HRSSSTRSGGDLTLGLEPSEEEAPRSL-----APSEGAGSDVFDGDLGMAAGLQSL 1101
Db 599 -NSPST-----SRTPLLSSLSATSNNSATNCID-----RNGQGH 631

Qy 1102 PTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNOQVVRPQPPSPREGPLPA 1159
Db 632 PVREDSFVQRYSSDPTGNFLEESIDDGFL-----PAPEYVNO--LMPKFS----- 675

Qy 1160 ARPAGATLERAKTLSPGKGVVWVDV-----AFGAVENPEYL 1197
Db 676 -----TAMVQNOIYNNISLTAISKLPMDSRYNQNSHSTAVDNPEYL 715

RESULT 11
Q86714 ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
```

```
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON-TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CB7ADF725E1 CRC64;

Query Match 25.2%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 4.4e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

Qy 578 GPEADQCACAHYKDPFCVACRSGVKPDLUSYMPIKFPDEGACQPCPINCTHSCVDL 637
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 1 GP--DHCWKCAHFIDGHCVCACAGVLGENDTL-VWKYADANAVCOLCHNCNCRGCKGP 57

Qy 638 DDKCPAEORASPLTSIVSAV-VGILLVVLGVVFGILIKERQCKIRKYTHRRLLQETEL 696
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 58 GLEGCP---NGSKTPSAAGVGLLCLVVGGLGIGLYLRRR-HIVRKRTLRRLQEREL 113

Qy 697 VEPLTPSGAMPNQAMRIKTELKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVL 756
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 114 VEPLTPSGEAPNQAHRLIKETEFKVKVVLGSGAGFTVYKGLWIPGEKVKIPVAIKEL 173

Qy 757 ENTSPKANKEILDEAYVWAGVSPVSRLLGICLTSTVOLVQLMPYGCCLLDHVRNRR 816
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 174 EATSPKANKEILDEAYVWASVDNPRVCRLLGICLTSTVOLITQLMPYGCCLLDYIREHKDN 233

Qy 817 LGSODLLNCWOIAKGSYLEDVRLVHRDLAARNVLKSPNNVKITDFGLARLLDIDE 876
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 234 IGSQVLLNCVOIAKGMNLYBEERLVRDLAARNVLKTPQHVKITDFGLAKLGADEKE 293

Qy 877 YHAGGKVPKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIREIPOLLE 936
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 294 YHAEKGKVPKMALESILRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPIASISSVLE 353

Qy 937 KGERLPOPICTIDVYIMVWKWIDSECRPRFRELVSFSEFMRDAPQRFVVIQ-NEDLG 995
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 354 KGERLPOPICTIDVYIMVWKWIDADSRPKFRELIAEFSEKARDPPRYLVIQDERMH 413

Qy 996 PASPLDSTFYRSLDEDDMGDLVDAEYLVPOQGFCCPDAPGAGVWVHRRSSSTRSG 1055
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 414 LPSPTDSKFTYLTNEEDMEDIVDAEYLVPHQGF-----NSPT--- 454

Qy 1056 GGDLTGLPSEEEAPRSL-----APSEGAGSVDFDGLGMAAKGLQSLTHDPSPQL 1110
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 455 -----SRTPLLSLSATSNSATNCID-----RNGQGHVPVEDSFVQ 491

Qy 1111 RYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNPQDVVRPQPPSPREGPLPAARAGATLE 1168
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 492 RYSSDPTGNFLEESIDGFL-----PAPEYVNO--LMPKKPS----- 526

Qy 1169 RAKTLSPCKNGVVKDF-----AFGGAIVENPEYL 1197
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 527 ----TAMVQNIYNNISLTAISKLPMSRYONSHSTAVDNPYL 566

Qy 1197 RAKTLSPCKNGVVKDF-----AFGGAIVENPEYL 1197
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 527 ----TAMVQNIYNNISLTAISKLPMSRYONSHSTAVDNPYL 566

RESULT 12
Q8WYV0
ID Q8WYV0 PRELIMINARY; PRT; 412 AA.
AC Q8WYV0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN P31659
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
```

```
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF18349; AAL55856.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP motif.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.9%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 9.6e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

Qy 889 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIREIPDLLEKGERLPQPICT 948
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 1 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIREIPDLLEKGERLPQPICT 60

Qy 949 IDVTIMVWKWIDSECRPRFRELVSFSEFMRDAPQRFVVIQ-NEDLG 1008
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 61 IDVTIMVWKWIDSECRPRFRELVSFSEFMRDAPQRFVVIQ-NEDLG 120

Qy 1009 LEDDMGDLVDAEYLVPOQGFCCPDAPGAGVWVHRRSSSTRSGGDLTLGLEPSEE 1068
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 121 LEDDMGDLVDAEYLVPOQGFCCPDAPGAGVWVHRRSSSTRSGGDLTLGLEPSEE 180

Qy 1069 EAPRSLPAPSEGAGSVDFDGLGMAAKGLQSLTHDPSPQRYSEDPTVPLPSETDGYV 1128
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 181 EAPRSLPAPSEGAGSVDFDGLGMAAKGLQSLTHDPSPQRYSEDPTVPLPSETDGYV 240

Qy 1129 APLTCSQPEYVNPQDVVRPQPPSPREGPLPAARAGATLERAKTLSPCKNGVVKDVPAFG 1188
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 241 APLTCSQPEYVNPQDVVRPQPPSPREGPLPAARAGATLERAKTLSPCKNGVVKDVPAFG 300

Qy 1189 GAVENPEYLTQGGAAPOP-----HPPPA---FSPAQDNL 1220
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 301 GAVENPEYLTQGGAAPOP-----HPPPA---FSPAQDNL 360

Qy 1221 YYWD-QDPPER-----GAPPSTFKTPTAEN 1245
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 361 WWTQCEPEGVRRSPDVSSGREGLTSAKIKRWEGPPTTSRGTCHARN 410

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
GN GAG, v-erb-A, v-erb-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RA Bruskina A., Jackson J., Bishop J.M., McCarty D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B."
RL Oncogene 5:15-24(1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -.

```

```
DR EMBL; X52211; CAA36459.1; JOINED.
DR HSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stdhrmn_receptor.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR001628; Znf_Casteroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000001; Euk_pkinase; 1.
DR ProDom; PD000035; Znf_Casteroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; TyrKc; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Nuclease; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MM; 3C5AED791E4E95CE CRC64;

Query Match 24.4%; Score 1662.5; DB 15; Length 962;
Best Local Similarity 39.7%; Pred. No. 1.3e-116;
Matches 420; Conservative 115; Mismatches 253; Indels 271; Gaps 37;

QY 248 KHSCLAC-----LHNHSGICELHCPALVTYNTDTFESMPNEGRTYFGASCVTACPY 301
DB 65 KDEQCVCGDKPTGYHYR-----C-ITCGCKSPFRITQKNLHP-----TYSCY 109

QY 302 NYLSTDVGSCTLVCLPHNQEVTAEDGTQCEKCS-KPCARVCYGLGMQYIKANSKFTGIT 360
DB 110 D-----GCCVIDKITRNQ-----CQLCRFKKICISV-CWAMDVLVDDSKRAKR 151

QY 361 EL-----EFAGCKKIFGSLAFPS-----FDGDPASNTAPQYIKANSKFI--- 401
DB 152 KLTEENRRRRKEEMIKSLQHRPSPSAEWEELIHVVTEAHRSTNAQSHWKQRKFLLED 211

QY 402 -----GITELGYL--VISAMPDSLPLDSVFQNLQVIRGRILHNG-- 439
DB 212 IGQSPHASMLDGDKVDLEATEFTFKIITPAITRVVDFAKNLPMSFELPCEDQIILLGCC 271

QY 440 -----AYSLLTQG-----LGISWTLGRS--LRELGSLALIH-HNTH 473
DB 272 MEIMSLRAAVRYDPESETTLTSGEMAVKREQLKNGGLGVVSDAIFDLGKLSAFNLDDE 331

QY 474 LCFVHTVPDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQVCNCSQ 533
DB 332 VALLQAV-----LLMSSDR-----TGLIC----- 350

QY 534 FLRQGEQCEVRCVLOGLPRE-YVNAR-HCLP----- 562
DB 351 -----VDKIEKQCSYLLAFPHYINRYKHNIPHFWSKLLMKVADLRMTGAHRSRFLHMV 406

QY 563 -CHPECOQNGSVTCFQPEADQCACAHYKDDPFCVACRPSGVKPDLSYMPIWKFPDDEG 621
DB 407 ECPTELSPOE-----VGP--DHCKMCAHFIDGPHCVKACPAVLGENDTL-VMKYADANA 458

QY 622 ACQPCPINTHSCVDLDDKCPAEQASPLTISVAVV-GILLVVLGVVFGILIKRQ 680
DB 459 VCQLCHPNCTRGCKPGLEGCP---NGSKTPSIAAGVVGGLCLLVVVGILGLYLRH-H 514

QY 681 KIRKYTWRLRLQETELVEPLTPSGAMPNQOMRILKETURKVKVLGSGAGFYVYGIWI 740
DB 515 IVRKRTLRLLQRELVPELTPGGEAPNQAHRLIKETEFKKVKVLGFGAFYVYKGLWI 574

QY 741 PGENYKIPVAIKVLRNTSPKANKELDEAYVMAGVSPYVSRLLGICLTSTVQLVTQL 800
DB 575 PEGEKVTIPVAIKELREATSPKANKELDEAYVMASVDNPHVCRLLGICLTSTVQLITQL 634
```

```
QY 801 MPYCLLDHVRENRLGSGQDLLNWCQIAKMGMSYLEVRLVHRDLAAARNLVKSPNHVK 860
DB 635 MPYCLLDYIREHKDNIGSQYLLNWCQIAKMGMYLEERHMHVHRDLAAARNLVKTPQHVK 694

QY 861 ITDGLARLLDIDETEHADCGKVPWKMALESILRRFTHQSDVWSYGVTVWELMTGA 920
DB 695 ITDFGLAKQLGADEKEYHAEGGKVPWKMALESILHRIYTHQSDVWSYGVTVWELMTGS 754

QY 921 KPYDGIIPAREIPDLLEGERLPQPPICITIDVYIMVVKCMMIDSECRPRFRELVSFSRMA 980
DB 755 KPYDGIIPASEISSVLEKGERLPQPPICITIDVYIMVVKCMMGAGSRPRFRELIAFSKMA 814

QY 981 RDPQRFVVIQ-NEDLGPAASPLDSTFYRSLLDDMDGLVDAAEYLVPOQGFCCPDAPCA 1039
DB 815 RDPRLVVIQGDERMHLPSPTDSKFYRTLMBEEDMEDIVDAEYLVPHQGFF----- 866

QY 1040 GGMVHHRHSSTRSGGDLTLGLEPSEEAAPRSPFLAPSEGAGSDVDFCDLGMGAAGLQ 1099
DB 867 -----NSPST-----SRTPLLSSLSATSN-----NSATKCID 893

QY 1100 SLPTHDPSPLOQRYSEDTVPPLPSETDGVVAPLTCSPOPEYVNPDPVRPQPPSPREGPLPA 1159
DB 894 RNCGH-----PVREDGFL-----PAPEYVNO--LMPKPFSTAMVQNOI 929

QY 1160 ARPAGAT-LERAKTLPSPKNGVKDVFAGFGAVENPEYL 1197
DB 930 YNYISLTAISKLPMSRYQN-----SHSTAVDNPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Te34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=28971102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278 (1987).
DR EMBL; X06943; CAA30024.1; -.
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 68899 MM; 140DCE8CCA0F8AF4 CRC64;

Query Match 24.1%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 1.3e-115;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 578 GPADQCVACAHYKDDPFCVACRPSGVKPDLSYMPIWKFPDEGACQPCPINTHSCVDL 637
DB 1 GP--DHCKMCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCQLCHPNCTRGCKGP 57

QY 638 DDGKCPAEQASPLTISVAVV-GILLVVLGVVFGILIKRQOKIRKYTWRLRLQETEL 696
DB 58 GLEGCP---NGSKTPSIAAGVVGGLCLLVVVGILGLYLRH-HIVKRTLRLLQREL 113
```

```
Qy 697 VEPLTPSGAMPNOAQRILKETELRKVKVLGSGAGFTYKGINIPDGENVKIPVAIKVL 756
Db 114 VEPLTPSGEAPNOAHLRLKETEPKVKVLGFGAGFTYKGLWIPEGEKVTIPVAIKELR 173
Qy 757 ENTSPKANKEITLDEAYVMAGVGSFYVSRLLGICLTSTVOLVQLMPYGCCLLDHVRENR 816
Db 174 EATSPKANKEITLDEAYVMASVDNPHVCRLLGICLTSTVOLVQLMPYGCCLLDYIREHKN 233
Qy 817 LGSQDLNWCWQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITIDFGLARLLDIDETE 876
Db 234 IGSQYLLNWCQIAKGNVLEERHLVHRDLAARNVLKTPQDVKITIDFGLAKOLGADEKE 293
Qy 877 YHADGGKVPKMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLE 936
Db 294 YHABGGKVPKMALESILRIYTHQSDVMSYGVTVWELMTFGSKPYDGPASEISSVLE 353
Qy 937 KGERLPOPPCTIDVYIMVWKMWIDSECRPFRELVSERMDARDQRFVVIQ-NEDLG 995
Db 354 KGERLPOPPCTIDVYIMVWKMWSDADSRPKFRELIAEFSKMDARDPPRYLVLIQDERMH 413
Qy 996 PASPLDSTFYRSLLEDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHHRHRSSTRSG 1055
Db 414 LPSPDTSKFTYTLNEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
Qy 1056 GGDLTGLPESEEAAPRSL-----APSEGAGSDVFDGDLGMGAAGLQSLPTHDPSPLO 1110
Db 455 -----SRTPLLSSLSATSNNSATNCIDRNGG-----H----- 481
Qy 1111 RYSEDTPVLPSETDGVYAPLTCSPQBYVNOPVDRPQPSREGPLPAAPAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVNO--LMPKKGSTAMVQIQIYNYISLTAISK 523
Qy 1170 AKTLPSPGNGVVKDVFAFGGAVENPEYL 1197
Db 524 LPWDSRYQN-----SHSTAVDNPEYL 544

RESULT 15
Q9MVF5 PRELIMINARY; PRT; 655 AA.
AC Q9MVF5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C., Lampland A.L., Balasubramanian S., Crossley T.O., Magnuson T.R., Maible N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7TAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Pearshall R.S., Green P.J., Yee D., Lampland A.L., Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D., Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egr transcripts encoding truncated receptor isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Iehii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Azawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S., Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:585-690(2001).
DR EMBL; AF124513; AAD44149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -.
DR EMBL; AK004893; BAB23641.1; -.
DR EMBL; AK004911; BAB23662.1; -.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; Fu; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;
```

```
Query Match 21.4%; Score 1458.5; DB 11; Length 655;
Best Local Similarity 43.8%; Pred. No. 2.1e-101;
Matches 280; Conservative 94; Mismatches 241; Indels 25; Gaps 9;

Qy 11 LLLALLPPGAA--STQVCTGTDMLKLPASPETHLDMLRHLQGVVQVGNLEITYPTN 68
Db 14 LTTALCAAGALBEKKVCCQGTNSRLTQLGTGFEDHFLSLQRMNCEVVLGNLEITYQRN 73
Qy 69 ASLSFLDIOIEVGVYLIANHQRVQVPLQRLRIVRGTLFEDNYALAVLDNGDPLNNTTP 128
Db 74 YDLSFLKTIQEVAGYVLIANTVETRIPLNQLIIRGNALYENTYVALAISN----- 124
Qy 129 VTGASPGLELOLRSLTEILKGVLIQRNPOLCYQDTILWKDI-----FHKNQLALTLI 184
Db 125 -YGNRTGLRELPMRLNQLIELIGAVRFSNPILCNMDTIQMRDIVQNVFMSNMDL--- 180
Qy 185 DTRNSRACHPCSPWCKSGRCWGESSEDCQSLTRTVCAAGCA-RCKGPLPTDCHEQCAAG 243
Db 181 -QSHPSCCPKCDSPCNGSCWGGEECNCKLTIKIAQCCSHRCGRSPDCCHNQCAAG 239
Qy 244 CTGPKHSDCLACLHFNHSGICELHCPALVYNTDTTFESMPNPEGRYTFGASCVTACPNY 303
Db 240 CTGPRESDCLVCKQFQDEATCKDTCPLMLNPTTYQMDVNPVEGKSFATCVKCKPRNY 299
Qy 304 LSTDVGSCLVCPHLHQEVTAEGTORCEKSPCARVCYGLGMOYIKANSKFGITELE 363
Db 300 VVTDHGSVCVRACGPDYVEV--EDGIRKCKCDGPRKVCNGIGIGEFK-DTSLINATNIK 357
Qy 364 -FAGCKKIFGSLAFLPESFDGDPASNTAPOYIKANSKFIGITELTGLVLYTSAPWDSPLDL 422
Db 358 HFYCYCTAISGDHLPLVAFKGDSTFTRPPLDPRELEILKTVKEITGELLTQAPDNWTDL 417
Qy 423 SVFQNLQVIRGRILHNGAVSLTLOGLGISWLGRLSRELGSGLALIHNTLHLCFVHTVPW 482
Db 418 HAFENLEIRGRTKHQHGFQSLAVVGLNITSLGRLSLKEISDGVVIISGNNLVCYANTINW 477
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-369-383-12

Perfect score: 6815
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6653	97.6	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5865	86.1	1257	1 ERB2_RAT	P06494 rattus norv
3	5851.5	85.9	1254	1 ERB2_MESAU	Q60553 mesocricetu
4	3095	45.4	1210	1 EGFR_HUMAN	P00533 homo sapien
5	3070	45.0	1210	1 EGFR_MOUSE	Q01279 mus musculu
6	2919.5	42.8	1308	1 ERB4_HUMAN	Q15303 homo sapien
7	2900	42.6	1308	1 ERB4_RAT	Q62956 rattus norv
8	2652.5	38.9	1167	1 XMRK_XIPWA	P13388 xiphophorus
9	2375.5	34.9	1342	1 ERB3_HUMAN	P21860 homo sapien
10	2303.5	33.8	1339	1 ERB3_RAT	Q62799 rattus norv
11	1922	28.2	1426	1 EGFR_DROME	P04412 drosophila
12	1749.5	25.7	634	1 ERBB_ALV	P00534 avian leuko
13	1703	25.0	604	1 ERBB_AVIER	P00535 avian eryth
14	1630	23.9	540	1 ERBB_AVIEU	P11273 avian eryth
15	1553	22.8	703	1 EGFR_CHICK	P13387 gallus gall
16	1275	18.7	1323	1 LT23_CAEL	P24348 caenorhabdi
17	1142.5	16.8	245	1 ERB2_MOUSE	P70424 mus musculu
18	730	10.7	1363	1 ILPR_BRALA	O02466 branchiosco
19	696	10.2	1383	1 INSR_RAT	P15127 rattus norv
20	696	10.2	1607	1 MIPR_LYMST	Q25410 lymnaea sca
21	695.5	10.2	1372	1 INSR_MOUSE	P15208 mus musculu
22	695	10.2	1382	1 INSR_HUMAN	P06213 homo sapien
23	689	10.1	1300	1 IRR_MOUSE	Q9W144 mus musculu
24	682	10.0	1297	1 IRR_HUMAN	P14616 homo sapien
25	676	9.9	1477	1 HTK7_HYDAT	Q25197 hydra atten
26	674.5	9.9	1300	1 IRR_CAVPO	P14617 cavia porce
27	644	9.4	1367	1 IGR_HUMAN	P08069 homo sapien
28	627	9.2	1390	1 INSR_AEDAE	Q93105 aedes aegypt
29	626	9.2	1373	1 IGR_MOUSE	Q60751 mus musculu
30	622.5	9.1	1370	1 IGR_RAT	P24062 rattus norv
31	612	9.0	2146	1 INSR_DROME	P09208 drosophila
32	599	8.8	987	1 EPB4_HUMAN	P54760 homo sapien
33	588	8.6	1114	1 RET_HUMAN	P07949 homo sapien

RESULT 1

ID	ERB2_HUMAN	STANDARD;	PRT;	1255 AA.
AC	P04626;			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)			
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell surface receptor HER2) (MLN 19).			
GN	ERBB2 OR HER2 OR NGL OR NEU.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86118663; PubMed=3003577;			
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,			
RA	Saito T., Toyoshima K.;			
RT	"Similarity of protein encoded by the human c-erb-B-2 gene to			
RT	epidermal growth factor receptor.";			
RL	Nature 319:230-234(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86070181; PubMed=2999974;			
RA	Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,			
RA	McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,			
RA	Frankke U., Levinson A., Ullrich A.;			
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor			
RT	shares chromosomal location with neu oncogene.";			
RL	Science 230:1132-1139(1985).			
RN	[3]			
RP	SEQUENCE OF 737-1031 FROM N.A.			
RX	MEDLINE=86016729; PubMed=2995967;			
RA	Semba K., Kamata N., Toyoshima K., Yamamoto T.;			
RT	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the			
RT	c-erbB-1/epidermal growth factor-receptor gene and is amplified in a			
RT	human salivary gland adenocarcinoma.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).			
RN	[4]			
RP	VARIANTs VAL-654 AND VAL-655.			
RX	MEDLINE=93194196; PubMed=8095488;			
RA	Ehsani A., Low J., Wallace R.B., Wu A.M.;			
RT	"Characterization of a new allele of the human ERBB2 gene by allele-			
RT	specific competition hybridization.";			
RL	Genomics 15:426-429(1993)			
CC	-!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,			
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A			
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-			
CC	ALPHA AND AMPHIREGULIN.			
CC	-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein			
CC	tyrosine phosphate.			
CC	-!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS			
CC	(POTENTIAL).			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			

ALIGNMENTS

34	587.5	8.6	984	1	EPB1_RAT	P09759 rattus norv
35	585.5	8.6	984	1	EPB1_CHICK	Q07494 gallus gall
36	581.5	8.5	977	1	EPB2_MOUSE	Q03145 mus musculu
37	581.5	8.5	984	1	EPB1_HUMAN	P54762 homo sapien
38	579	8.5	976	1	EPB2_HUMAN	P29317 homo sapien
39	576	8.5	902	1	EPBB_XENLA	Q91736 xenopus lae
40	573.5	8.4	1053	1	FAK1_CHICK	Q00944 gallus gall
41	570	8.4	987	1	EPB4_MOUSE	Q54761 mus musculu
42	569.5	8.4	985	1	EPBA_XENLA	Q91571 xenopus lae
43	569	8.3	757	1	HT16_HYDAT	P53356 hydra atten
44	569	8.3	1068	1	FAK1_XENLA	Q91738 xenopus lae
45	563	8.3	1052	1	FAK1_MOUSE	P34152 mus musculu

-!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
RESIDUES (BY SIMILARITY).
-!- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
POSITIONS 654 AND 655. ALLELE B1 (654-IIE-655) HAS A FREQUENCY
OF 0.782; ALLELE B2 (654-IIE-VAL-655) HAS A FREQUENCY OF 0.206;
ALLELE B3 (654-VAL-655) HAS A FREQUENCY OF 0.012.
-!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; M11767; AAA35808.1; -;
DR EMBL; M11761; AAA35808.1; JOINED.
DR EMBL; M11762; AAA35808.1; JOINED.
DR EMBL; M11763; AAA35808.1; JOINED.
DR EMBL; M11764; AAA35808.1; JOINED.
DR EMBL; M11765; AAA35808.1; JOINED.
DR EMBL; M11766; AAA35808.1; JOINED.
DR EMBL; M11730; AAA75493.1; -;
DR EMBL; M12036; AAA35978.1; -;
DR EMBL; X03363; CAA27060.1; -;
DR PIR; A25491; A25491.
DR PIR; A24571; A24571.
DR HSSP; P11362; 1FGK.
DR GenBank; HGNC:3430; ERBB2.
DR MIM; 164870; -;
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Polymorphism.
FT SIGNAL 1 21
FT CHAIN 22 1255
FT DOMAIN 22 652
FT TRANSMEM 653 675
FT DOMAIN 676 1255
FT DOMAIN 720 987
FT NP_BIND 726 734
FT BINDING 753 753
FT ACT_SITE 845 845
FT DISULFID 195 204
FT DISULFID 199 212
FT DISULFID 220 227
FT DISULFID 224 235
FT DISULFID 236 244
FT DISULFID 240 252
FT DISULFID 255 264
FT DISULFID 268 295
FT DISULFID 299 311
FT DISULFID 315 331
FT DISULFID 334 338
FT DISULFID 511 520
FT DISULFID 515 528
FT DISULFID 531 540
FT DISULFID 544 560

FT DISULFID 563 576
FT DISULFID 584 584
FT DISULFID 596 596
FT DISULFID 600 623
FT DISULFID 626 634
FT DISULFID 630 642
FT MOD_RES 1139 1139
FT MOD_RES 1248 1248
FT CARBOHYD 68 68
FT CARBOHYD 124 124
FT CARBOHYD 187 187
FT CARBOHYD 259 259
FT CARBOHYD 530 530
FT CARBOHYD 571 571
FT CARBOHYD 629 629
FT VARIANT 654 654
FT VARIANT 655 655
FT CONFLICT 1170 1170
FT SEQUENCE 1255 AA; 137909 MM; 39E9DFDA04DCF962 CRC64;
SQ
Query Match 97.6%; Score 6653; DB 1; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPPGAASQTCTGTDMLRLPASPTHLDMLRHLYQSCVQVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASQTCTGTDMLRLPASPTHLDMLRHLYQSCVQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIHQNVRQVPLQRLRIVRGTFQLEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIHQNVRQVPLQRLRIVRGTFQLEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
DB 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPCKRCWGESSEDCQSILTRTVCGAGCARCKGPLPTDCCHQEC 240
DB 181 LTLIDTNRSRACHPCSPCKRCWGESSEDCQSILTRTVCGAGCARCKGPLPTDCCHQEC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHNLQEVTAEDGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360
DB 301 YNYLSTDVGSCTLVCPHNLQEVTAEDGTQRCCKSKPCARVCYGLGMHLEVRVAVTSAN 360
QY 361 ELEFAGCKKITGSLAFLPESFDGDPASNTAPQYIKANSKFIGITELTGVLYISAMPDLSLP 420
DB 361 IQEFAGCKKITGSLAFLPESFDGDPASNTAPLQPEQLVFTLEBITGVLYISAMPDLSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLIGISWGLGRLSRLRELGSGLALIHNNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLIGISWGLGRLSRLRELGSGLALIHNNTHLCFVHTV 480
QY 481 PWDQLFRNPHOALLHTANRPEDECVGEGLAHQICARGHCWGPGTQCVNCSQFIRGQEC 540
DB 481 PWDQLFRNPHOALLHTANRPEDECVGEGLAHQICARGHCWGPGTQCVNCSQFIRGQEC 540
QY 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPPEFCVARC 600
DB 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPPEFCVARC 600
QY 601 PSGVKPDLSTYMPIWKFFDEEGACQPCPINCTHSCVDLDDKCGKPAEQRASPLTISIVAVVG 660
DB 601 PSGVKPDLSTYMPIWKFFDEEGACQPCPINCTHSCVDLDDKCGKPAEQRASPLTISIVAVVG 660
QY 661 ILLVVVLGVVFGIILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
DB 661 ILLVVVLGVVFGIILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy 721 RKVKVLSGARGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILLDEAYVMAGVGP 780
 Db 721 RKVKVLSGARGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILLDEAYVMAGVGP 780
 Qy 781 YVSRLLGLICTSTVQLVTQMLPYGCLLDVRENRGRGSLQLLNWCMIAGKMSYLEVDV 840
 Db 781 YVSRLLGLICTSTVQLVTQMLPYGCLLDVRENRGRGSLQLLNWCMIAGKMSYLEVDV 840
 Qy 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMKMALESILRRRFT 900
 Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMKMALESILRRRFT 900
 Qy 901 HOSDVWSVGVTVWELMTFGAKPYDGPAREIPDILEKGERLPPOPICTIDVYIMVKWM 960
 Db 901 HOSDVWSVGVTVWELMTFGAKPYDGPAREIPDILEKGERLPPOPICTIDVYIMVKWM 960
 Qy 961 IDSECRPRFRELVSFEFSMARDPQRFVVIQNEGLGPASPLDSTFYRSLLDDMDGLVDA 1020
 Db 961 IDSECRPRFRELVSFEFSMARDPQRFVVIQNEGLGPASPLDSTFYRSLLDDMDGLVDA 1020
 Qy 1021 EYILVPOQGFPCPDAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 Db 1021 EYILVPOQGFPCPDAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 Qy 1081 AGSDVFDGLGKGAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQBPYV 1140
 Db 1081 AGSDVFDGLGKGAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQBPYV 1140
 Qy 1141 NQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKGVVDVFAFGGAVENPEYLTQ 1200
 Db 1141 NQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKGVVDVFAFGGAVENPEYLTQ 1200
 Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPAENPEYLGLDVFPV 1255
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPAENPEYLGLDVFPV 1255

RESULT 2
 ERB2 RAT STANDARD; PRT; 1257 AA.
 AC P06494;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
 DE receptor-related protein).
 GN ERB2 OR NEU.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxId=10116;
 RN (1)
 RN SEQUENCE FROM N.A.
 RC TISSUE=Neuroblastoma;
 RX MEDLINE=86118662; PubMed=3945311;
 RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
 RT "The neu oncogene encodes an epidermal growth factor receptor-related
 RT protein.";
 RL Nature 319:226-230 (1986).
 RN (2)
 RP SEQUENCE OF 852-905 FROM N.A.
 RC TISSUE=Sciatic nerve;
 RX MEDLINE=91222560; PubMed=2025425;
 RA Lai C., Lemke G.;
 RT "An extended family of protein-tyrosine kinase genes differentially
 RT expressed in the vertebrate nervous system.";
 RL Neuron 6:691-704 (1991).
 RN (3)
 RP STRUCTURE BY NMR OF 650-668.
 RX MEDLINE=92155181; PubMed=1346763;
 RA Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D.,

RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
 RT "Three dimensional structure of the transmembrane region of the proto-
 RT oncogenic and oncogenic forms of the neu protein.";
 RL EMBO J. 11:43-48 (1992).
 CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP10 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
 CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- PM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X03362; CAA27059.1; ALT_INIT.
 DR PIR; A24562; TVRTNU.
 DR HSP; P11362; 1FGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR InterPro; IPR004019; YLP motif.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 655 677 POTENTIAL.
 FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 159 369 CYS-RICH.
 FT DOMAIN 473 646 CYS-RICH.
 FT DOMAIN 722 989 PROTEIN KINASE.
 FT NP_BIND 728 736 ATP (BY SIMILARITY).
 FT BINDING 755 755 ATP (BY SIMILARITY).
 FT ACT_SITE 847 847 BY SIMILARITY.
 FT DISULFID 196 205 BY SIMILARITY.
 FT DISULFID 200 213 BY SIMILARITY.
 FT DISULFID 221 228 BY SIMILARITY.
 FT DISULFID 225 236 BY SIMILARITY.
 FT DISULFID 237 245 BY SIMILARITY.
 FT DISULFID 241 253 BY SIMILARITY.
 FT DISULFID 256 265 BY SIMILARITY.
 FT DISULFID 269 296 BY SIMILARITY.
 FT DISULFID 300 312 BY SIMILARITY.
 FT DISULFID 316 332 BY SIMILARITY.
 FT DISULFID 335 339 BY SIMILARITY.
 FT DISULFID 513 522 BY SIMILARITY.
 FT DISULFID 517 530 BY SIMILARITY.
 FT DISULFID 533 542 BY SIMILARITY.
 FT DISULFID 546 562 BY SIMILARITY.
 FT DISULFID 565 578 BY SIMILARITY.

FT	DISULFID	569	586	BY SIMILARITY.
FT	DISULFID	589	598	BY SIMILARITY.
FT	DISULFID	602	625	BY SIMILARITY.
FT	DISULFID	628	636	BY SIMILARITY.
FT	DISULFID	632	644	BY SIMILARITY.
FT	MOD. RES	1141	1141	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD. RES	1250	1250	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	68	68	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	188	188	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	260	260	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	532	532	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	573	573	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	631	631	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARIANT	661	661	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARIANT	661	661	V -> E (IN ONCOGENIC NEU).
SEQ	SEQUENCE	1257 AA;	138831 MW;	6129264583011402 CRC64;
Query Match				
Best Local Similarity 86.1%; Score 5865; DB 1; Length 1257;				
Matches 1082; Conservative 54; Mismatches 119; Indels 2; Gaps 2;				
Qy	1	MELAAALCRWGLLLALLPFGAAS	TOVCTGCTDMKRLRPASPETHLDMRLHLYOGCQVQGNL	60
Db	1	MELAAACRWGLLLALLPFGIAGT	QVCTGCTDMKRLRPASPETHLDMRLHLYOGCQVQGNL	60
Qy	61	ELTYLPNVAISFLQDIOEVGYV	LIAHQVQVPLQRLRIVRGQLFEDNYALAVLDNG	120
Db	61	ELTYVPANASLFLQDIOEVGYM	LIAHQVQVPLQRLRIVRGQLFEDKYALAVLDNR	120
Qy	121	DPLNNTTPTV	-GASPGGLREQLRLSLTEILKGGVLIORNPOLCYODTILKWDI	179
Db	121	DQONVAASTPGRTEGURELQ	RLSLTEILKGGVLIORNPOLCYODMWLVKDFRKNQL	180
Qy	180	ALTLDITNRSRACHPCSPMKG	SCWGSSEDCOSLRTVCAGGCARCKGLPTDCCHEQ	239
Db	181	APVDITNRSRACPPACAPACK	NDHCWGESPEDCQILTGICTSGCARCKGLPTDCCHEQ	240
Qy	240	CAAGCTGPKHSDCLACHFNH	SGICELHCPALVYNTDTFESMNPPEGRTYFGASCVTAC	299
Db	241	CAAGCTGPKHSDCLACHFNH	SGICELHCPALVYNTDTFESMNPPEGRTYFGASCVTTC	300
Qy	300	PYNLYSTDVSGCTVLCVPLH	NOEVTAEDGTQCEKSCPCARVCYGLGQYIKANSKFIGI	359
Db	301	PYNLYSTEVGCTVLCVPPN	NOEVTAEDGTQCEKSCPCARVCYGLGMEHLRGARATSD	360
Qy	360	TELEFAGCKIFGSLAFPE	SGDGPASNTAPQVIKANSKFIGITELTYLYISAWPDSL	419
Db	361	NVQEFDGCKIFGSLAFPE	SGDGPSSGIAPLRPEQLQVFETLEEITGYLYISAWPDSL	420
Qy	420	PDLVSFQNLQVIRGILHNG	AYSULTQGLGHSWGLSLRLSLRGLSGLALIHNTLCFVHT	479
Db	421	RDLVSFQNLRIIRGILHDG	AYSULTQGLGHSWGLSLRLSLRGLSGLALIHNTLCFVHT	480
Qy	480	VPWDLFRNPHOALLHTANR	PEDE-CVGEGLACHOLCARGHCWGPPTQCVNCSQFLRGQ	538
Db	481	VPWDLFRNPHOALLHNSNR	PEDELCVSSGLVNCSLCHGHCWGPPTQCVNCSHFRLRGQ	540
Qy	539	ECVEECRVQLPLREYNARH	CLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVA	598
Db	541	ECVEECRVKGLPREYVSDK	RLCPCHPECPQNSSETCFGEADQCAACAHYKDSSCVA	600
Qy	599	RCPSGVKPDLSYMPDKFPE	DEGACQPCPINCTHSCVDLDDKGCAPACORASPLTSISAV	658
Db	601	RCPSGVKPDLSYMPDKFPE	DEGICQPCPINCTHSCVDLDERGCPAECRASPVTFIATV	660
Qy	659	VGILLVVVLGVVFGILIKR	ROOKIRKYTMRLLOTELVEPLTPSGAMPNOAQRILKET	718
Db	661	VGILLVLIVVVGILIKRR	QKIRKYTMRLLOTELVEPLTPSGAMPNOAQRILKET	720
Qy	719	ELRKVKVLGSAFTVYKGIW	IPDGENYKIPVAIKVIRENTSPKANKEILDEAYVMAGV	778
Db	721	ELRKVKVLGSAFTVYKGIW	IPDGENYKIPVAIKVIRENTSPKANKEILDEAYVMAGV	780
Qy	779	SPYVSRLLGICLTSTVQL	VTQMLPYPGCLLDHVRNRRGRGLSQDILLNWCMIAGKMSYLED	838

Db	781	SPVSRLLGICLTSTVQLVTQMLPYPGCLLDHVRHRRGRIGSQDILLNWCVOIAGKMSYLED	840	
Qy	839	VRLVHRDLAARNVLVKSPNHVKITDFGLARLLDDIDETEHADGGKVPICKMALESILRRR	898	
Db	841	VRLVHRDLAARNVLVKSPNHVKITDFGLARLLDDIDETEHADGGKVPICKMALESILRRR	900	
Qy	899	FTTHQSDVMSYGVTVWELMTFCAPYDGPAREIPDLLEKGERLPOPPICITDVMYIMVKC	958	
Db	901	FTTHQSDVMSYGVTVWELMTFCAPYDGPAREIPDLLEKGERLPOPPICITDVMYIMVKC	960	
Qy	959	WMIDSECRPRFRELVSFESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDMGDLV	1018	
Db	961	WMIDSECRPRFRELVSFESRMARDPQRFVVIQNEDLGPSSPMDSTFYRSLLEDDDDMGDLV	1020	
Qy	1019	DAEYLVLPQGGFFCPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAAPRSPAPS	1078	
Db	1021	DAEYLVLPQGGFFSPDPTPGTGTAHRRSSSTRSGGDLTLGLEPSEEGPSPSPAPS	1080	
Qy	1079	EGAGSDVDFDGLGMAAKGLQLSPHTDPSPLQRYSEDPVLPSETDGYVAPLTCSPQPE	1138	
Db	1081	EGAGSDVDFDGLGMAAGVTGKLOSLSPHTDPSPLQRYSEDPVLPSETDGYVAPLTCSPQPE	1140	
Qy	1139	YVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLT	1198	
Db	1141	YVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLV	1200	
Qy	1199	POGGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPPTAENPEYLGLDVVPV	1255	
Db	1201	PREGTAPPHSPAFSPAFDNLVYWDQNSSEQGPPSPNFEPTPTAENPEYLGLDVVPV	1257	
RESULT 3				
ERB2_MESAU				
ID	ERB2_MESAU	STANDARD;	PRT;	1254 AA.
AC	Q60553;			
DT	15-DEC-1998	(Rel. 37, Created)		
DT	15-DEC-1998	(Rel. 37, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)			
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2).			
GN	ERB2 OR NEU.			
OS	Mesocricetus auratus (Golden hamster).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;			
OC	Mesocricetus.			
OX	NCBI_TaxID=10036;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	TISSUE=Nerve;			
RX	MEDLINE=94191007; PubMed=7908275;			
RA	Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,			
RA	Yamazaki Y., Ishikawa T.;			
RT	"Cloning and activation of the Syrian hamster neu proto-oncogene.";			
RT	Gene 140:251-255(1994).			
CC	-!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,			
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A			
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-			
CC	ALPHA AND AMPHIREGULIN (BY SIMILARITY).			
CC	-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein			
CC	tyrosine phosphate.			
CC	-!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS			
CC	(POTENTIAL).			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE			
CC	RESIDUES.			
CC	-!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.			

CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC	EMBL; D16295; BAA03801.1; --	
DR	HSRP; P11362; IFGK.	
DR	InterPro; IPR000494; EGFR_L_domain.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001245; Tyr_pkinase.	
DR	InterPro; IPR004019; YLP_motif.	
DR	Pfam; PF00089; pkinase; 1.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF01030; Recep_L_domain; 2.	
DR	Pfam; PF02757; YLP; 2.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00361; FU; 3.	
DR	SMART; SM00319; TyKc; 1.	
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.	
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.	
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	
KW	Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;	
KW	Proto-oncogene; Disease mutation.	
FT	SIGNAL 1 21	POTENTIAL.
FT	CHAIN 22 1254	RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT	DOMAIN 22 652	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 653 675	POTENTIAL.
FT	DOMAIN 676 1254	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 158 368	CYS-RICH.
FT	DOMAIN 472 644	CYS-RICH.
FT	DOMAIN 720 987	PROTEIN KINASE.
FT	NP_BIND 726 734	ATP (BY SIMILARITY).
FT	BINDING 753 753	ATP (BY SIMILARITY).
FT	ACT_SITE 845 845	BY SIMILARITY.
FT	DISULFID 195 204	BY SIMILARITY.
FT	DISULFID 199 212	BY SIMILARITY.
FT	DISULFID 236 244	BY SIMILARITY.
FT	DISULFID 240 252	BY SIMILARITY.
FT	DISULFID 255 264	BY SIMILARITY.
FT	DISULFID 268 295	BY SIMILARITY.
FT	DISULFID 299 311	BY SIMILARITY.
FT	DISULFID 315 331	BY SIMILARITY.
FT	DISULFID 334 338	BY SIMILARITY.
FT	DISULFID 511 520	BY SIMILARITY.
FT	DISULFID 515 528	BY SIMILARITY.
FT	DISULFID 531 540	BY SIMILARITY.
FT	DISULFID 544 560	BY SIMILARITY.
FT	DISULFID 563 576	BY SIMILARITY.
FT	DISULFID 567 584	BY SIMILARITY.
FT	DISULFID 587 596	BY SIMILARITY.
FT	DISULFID 600 623	BY SIMILARITY.
FT	DISULFID 626 634	BY SIMILARITY.
FT	DISULFID 630 642	BY SIMILARITY.
FT	MOD_RES 1139 1139	BY SIMILARITY.
FT	MOD_RES 1247 1247	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD 68 68	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 125 125	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 187 187	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 259 259	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 530 530	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 571 571	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 629 629	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT 658 658	V -> E (IN ONCOGENIC NEU).
FT	VARIANT 659 659	V -> E (IN ONCOGENIC NEU).
FT	SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2B81 CRC64;	
CC		

Qy	61	ELTYLPTNASLFLQDIOEVQGYVYLIAHNOVRQVPLQRLRIVRGTLQFEDNYALAVLQNG	120
Dd	61	ELTYLPANATLSFLQDIOEVQGYWLIIAHSQVRHVPLOQLRIVRGTLQFEDYKALAVLNR	120
Qy	121	DPLNNTTFTVGTASPGGLRELQLRSLEILKGGVLIQIRNPQLCYQDTILWKDIFHKNNOLA	180
Dd	121	DPLDNVTATGRTPEGLRELQLRSLEILKGGVLIIRGNPQLCYQDTVLWKDVFKNQOLA	180
Qy	181	LTLIDTNRBRACHPCSPMKSGSRCWGSSBDCQSLRTRTVCAGGCARCKGPIPTDCCBQC	240
Dd	181	PVDIDTNRBRACPPCAPACKDNHCWGASPEDCQTLTGTIAPRAVPAARLARLPDTCBQC	240
Qy	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Dd	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTCP	300
Qy	301	YNYLSTDVSGCTLVCPLHNOEVTAEQDTCRCEKSKPCARVCYGLGMOYIKANSKFIGIT	360
Dd	301	YNYLSTEVGSCTLVCLPNNQEVTAEDTQRCCKSKSCARVCYGLGMEHLRGARAITSAN	360
Qy	361	ELEPAGCKKIFGSLAFIPESFGDGPASNTAPQYIKANSKFITGILTGLYVISAWPDSL	420
Dd	361	IQBPAGCKKIFGSLAFIPESFGDGNPSSGIAPLTPEQLQVFPTELEEITGILYVISAWPDSL	420
Qy	421	DLVSFQNLQVIRGRILHNGAYSITLOGIGISWLGRLSRLRELSGSLALIHNNTHLCFVHTV	480
Dd	421	DLVSFQNLQVIRGRVLHDGAYSIALOGLGIRWLGLSRLRELSGSLALVIRHNNTHLCFVHTV	480
Qy	481	PWDLFRNPQALLHTANRPEDEKCVGSLGACHOLCARGHCGWGPCTOCVNCQSOFLRQEC	540
Dd	481	PWDLFRNPQALLHSGNPSSEECGLKDFACYPICAHGHCWGPPTQCVNCSHFLRGQEC	540
Qy	541	VEBCRVLQGLPREYVNNARHCLPCHPECOQPNQSVTCFQPEADQCAAHYKDPDFCVARC	600
Dd	541	VKECRVWKGLPREYVNGKHCLPCHPECOQPNSTETCTGSEADQCTAPHYKDSFFCVARC	600
Qy	601	PSGVKPDLSYMPITWKFDPBEGACQCPDINCTHSCVDLDDKGCQPAEQRASPUTSIVSAVVG	660
Dd	601	PSGVKPDLSYMPITWKYPDEEGMCQPCPINCTHSCVDLDERGCPAQEQRASPATSIATVVG	660
Qy	661	ILLVVLGVVFGILIKRQOKIRKTYMRRLLQETELVEPLTPSGAMPNQOMRILKETEL	720
Dd	661	ILLFLVIGVVVGILIKRRQOKIRKTYMRRLLQETELVEPLTPSGAMPNQOMRILKETEL	720
Qy	721	RKVKVLGSGAFGYVKGITPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVSP	780
Dd	721	RKVKVLGSGAFGYVKGITPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGLSP	780
Qy	781	YVSRLLGICLUTSVQLVTQMPYGCLLDHVRENRGRLSQDLLNWCQIAKGMYSYLEDVR	840
Dd	781	YVSRLLGICLUTSVQLVTQMPYGCLLDHVREHRGRLSQDLLNWCQVIAKGMYSYLEDVR	840
Qy	841	LVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETETHADGGKVPKIKWMALESILRRRFT	900
Dd	841	LVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETETHADGGKVPKIKWIALESILRRRFT	900
Qy	901	HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYIMVKCWM	960
Dd	901	HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYIMVKCWM	960
Qy	961	IDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDA	1020
Dd	961	IDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLEDDDMGDLVDA	1020
Qy	1021	EYVLVPQQGFFCDDPAPGAGGMVHRRHSSTSSRSGGDLTLGLEPSEEEAPRSPAPSEG	1080
Dd	1021	EYVLVPQQGFFPDPAAGAGSTAHRRHSSTSSRSGGELTLGMEPSEEPSPRSPAPSEG	1080
Qy	1081	AGSDVFDGDLGAAKGLQSLPHTDPSPLQRYSEDTVPPLPSETDGVVAPLTCSPQPEYV	1140
Dd	1081	AGSDVFEGLMGMATKGPQISPRDLSPLQRYSEDTPLPLPTETDGVVAPLACSPQPEYV	1140

OY 1141 NOPDVPPQPPSPREGPLPAARPPAGATLERAKTLSPGKNGVYKOVFAFGGAVENPEYLTPQ 1200
 DB 1141 NQPEVRPPQLTPEGLPPVRPAGATLERPKTLSPGKNGVYKOVFTFGGAVENPEYLVRP 1200
 OY 1201 GGAAPQHPHPPSPAFDNLNLYWDQPPERCAPPSTFGTPTAENPEYLGLDVPV 1255
 DB 1201 GGSASOPH-PPALCPAFDNLNLYWDQPPSERGSPPTFGTPTAENPEYLGLDVPV 1254

RESULT 4
 ID_EGFR_HUMAN STANDARD; PRT; 1210 AA.
 AC P00533; P06268; Q14225; Q9UMD8; Q9UMGS; Q92795; O00732;
 AC O00688; Q9B2S2; Q9H2C9; Q9GZX1; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 DE protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Varden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Hayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RL Nature 309:418-425 (1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=95382957; PubMed=7654368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta.";
 RL Mol. Reprod. Dev. 41:149-156 (1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=97078686; PubMed=8918811;
 RA Reiter J.L., Maible N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor.";
 RL Nucleic Acids Res. 24:4050-4056 (1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=97256547; PubMed=9103388;
 RA Ilekis J.V., Garti J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TEGFR) in ovarian cancer.";
 RL Gynecol. Oncol. 65:36-41 (1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RC TISSUE=Placenta;
 RX MEDLINE=21100872; PubMed=11161793;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Schenl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
 RA Maible N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms.";
 RL Genomics 71:1-20 (2001).
 RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,

RA Maible N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
 RT "Expression cloning of human EGF receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells.";
 RL Science 224:843-848 (1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells.";
 RL Nature 309:806-810 (1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
 RA O'Malley B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells.";
 RL Biochem. Biophys. Res. Commun. 124:125-132 (1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
 RA Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription.";
 RL Oncogene Res. 1:375-396 (1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RT factor receptor proto-oncogene RNA synthesis.";
 RL J. Biol. Chem. 266:1746-1753 (1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:4920-4924 (1985).
 RN [13]
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RL Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor
 RT and supercoiled DNA.";
 RL Nature 309:270-273 (1984).
 RN [15]
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor.";
 RL J. Biol. Chem. 264:10667-10671 (1989).
 RN [16]

FT	DISULFID	308	323	BY SIMILARITY.
FT	DISULFID	326	330	BY SIMILARITY.
FT	DISULFID	503	512	BY SIMILARITY.
FT	DISULFID	507	520	BY SIMILARITY.
FT	DISULFID	523	532	BY SIMILARITY.
FT	DISULFID	536	552	BY SIMILARITY.
FT	DISULFID	555	569	BY SIMILARITY.
FT	DISULFID	580	589	BY SIMILARITY.
FT	DISULFID	593	614	BY SIMILARITY.
FT	DISULFID	617	625	BY SIMILARITY.
FT	DISULFID	621	633	BY SIMILARITY.
FT	MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	174	174	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	181	181	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	253	253	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	358	358	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	410	410	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	473	473	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	495	495	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	548	548	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	620	620	N-LINKED (GLCNAC) (POTENTIAL).
FT	VARSPLIC	626	648	NGTSHDCIYPMWHTGSHTLPPQHA -> IGSSIEDICGLMD
FT				(IN ISOFORM JM-B)
SEQ	SEQUENCE	1308 AA;	146807 MW;	584A880985D88761 CRC64;
Query Match				
Best Local Similarity 42.8%; Score 2919.5; DB 1; Length 1308;				
Matches 603; Conservative 185; Mismatches 385; Indels 173; Gaps 28;				
Qy	9	WGILLALLPGAA-----	STQVCTGTDKMLRPLASPETHLDMLRHLVGGCVOVQGNLELT	64
Db	8	WVWSLLVAAGTVQPSDSQSVCAGTENKLSLSDLSEQQYRALRKYYENCEVMGNLEITS	67	
Qy	65	LPTNASLFLDIOEQVGYVLIAHNOVROVPLQRLRI	VRGTOLFDENYALAVLDNGDPLN	124
Db	68	IEHRDLSFLRSVREVTGVVVALANFRVLPLENLRI	IRGTKUYEDRYALAIFFNRYKOG	127
Qy	125	NTTPVTGASPGGLRELQLRSLTEILKGGVLI	IQRNPOLCYQDTILWKDIFHKNQLALTLI	184
Db	128	NF-----GLQELGLKMLTEILNGVGVVDQNKELCYADTIHWODIVRNPWPSNLTIV	178	
Qy	185	DTNRSRACHIPCSMPKSGRCWGSESSDCQSLRTRTVCAAGC-ARCKGPLPTDCHEQCAAG	243	
Db	179	STNGSSGCGCRCHKSCG-RWMPGTENHCOTLTRTVCAEQDCGRCYGYPVSDCCHRECAGG	237	
Qy	244	CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVYACPYNY	303	
Db	238	CSPGKDTDFACMWFNDSGACVTCQCPQTFVYNPPTTFQLEHFNAKYTFGAFVKKCPHNF	297	
Qy	304	LSTDVGSCTLVCPHLNHQEVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIGITELE	363	
Db	298	V-VDSSSVCRACPSSKMEV-EEINGIRKCKPCTDIPCACADGIGTGLMSAQTVDDSSNIDK	355	
Qy	364	PACCKKIFGSLAFPLPESFGDDPASNTAPQVIKANSKFIGITELTGYLYISAMPDLSIDLS	423	
Db	356	FINCTKINGNLIFLVGIGHGDPYNAIEADPEKLNIVFRTVREITGTFGLNQSWPPNMTDFS	415	
Qy	424	VFONLQVIRGRIIHNCAYSILTIQGLGISWGLGRSLRELCSGLALIHNNTHLCFVHTVPMD	483	
Db	416	VFSNLTIVIGRVYISGLSLLILKQQGITSLOQSLKEISAGNIYITDNSNLCCYHTINMT	475	
Qy	484	QLFRNPQALLHTANRPEDECVGEGLACHQLCARGHCWGPQTQCVCNCSQFRLRGQCEVE	543	
Db	476	TLEFSTINQIRIVRDNRKAENCTAEGMVCNHLCSDDCGCWGFGDOCLSCRRFRSGRICIES	535	
Qy	544	CRVLQGLPREYVNAHCLPCHPECPQ-QNGSVTCFGPEADQCVCACHYKDPFPFCVARCPES	602	


```
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE
FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;

Query Match 42.6%; Score 2900; DB 1; Length 1308;
Best Local Similarity 44.6%; Pred. No. 6.8e-147;
Matches 601; Conservative 192; Mismatches 394; Indels 160; Gaps 28;

Qy 1 MELA-ALCRWGLLL--ALLPPGAASQVCTGTDMLRLPASPETHLDMRLHLYQCQVQ 57
Db 1 MKLATGLVWWSLLVAARTVQPSASQSVCACTENKLSLSLEQQYRALRYENCEVVM 60

Qy 58 GNLELTYPNASHLSFLQDIOEQVYLIANOVROVPLQRLIRVGTQOLFEDNYALVL 117
Db 61 GNLEITSIEHRNDSLSFIRSIEVGYVLVALNQRYLPLENRLIRGTGLYEDRYALAIF 120

Qy 118 DNGDPLNNTTPTVGTASPGGLRELOLSLTKGLGVLIQRPOLCYQDTILWKDIEHKN 177
Db 121 LNYRKDGNF-----GLQELGLKNLTIELNGVGYVDQNKFLCYADTIHQDILVRNP 171

Qy 178 QLAULTIDNRSRACHPCSPMKGRWGSESDCQSLTRTVCAAGC-ARCKGLPTDCC 236
Db 172 PSNMTLVSTIGSSGCGRCHKSCGTG-RCWGPTENHCOTLTRVCAEQCDRCYGPVSDCC 230

Qy 237 HEQCAAGCTGPKHSDCLACLFNHSIGELCPALVTNTDTFSPMPNPEGRYTFGASCV 296
Db 231 HRECAAGCGSKPDTCDFACMNFNDSGACVTCQPTFVYNPTTTFQLEHNFNAKYTGAFCV 290

Qy 297 TACPYNYSLTGVGCTLLVCLPHNDEVTADGTQRCCKSKPCARVCVYGLGMYIKANSKF 356
Db 291 KKCCHNFV-VDSSSCVACAPSSKMEV-EENGIKMKCPCTDICPKACDGIPTGSLMSAQTV 348
```

RESULT 8
XMRK_XIPMA
ID XMRK_XIPMA
AC P13386;

STANDARD; PRT; 1167 AA.

```
Qy 357 IGITELBFAGCKKIFGSLAFLPESFDGDPASNTAPOVIKANSKFIGITELTGILYVISAMP 416
Db 349 DSSNIDKFINCTKINGNLIFLVGTGHGDPYNAIDAIDPEKLVNRTVREITGFLNIOTMP 408

Qy 417 DSLPDLVSFQNLQVIRGIRLHNGAYSITLQGLGTSWGLRSLRELGSGLAIHNNTHLCF 476
Db 409 PNMTDFSVFSLVITIGRVLYSGLSLILKQOGITSLQFQSLKEISAGNIYITONSILCY 468

Qy 477 VHTVPDOLERNPHOALLHTANRPEDECVEGLACHOLCARGHCWGCPGTPCVCNSQFLR 536
Db 469 YHTINWTLTSTVNRIVIRDRRAENCTAGMVCNHLCSNDGCMGPGPDQCLSCRRFSR 528

Qy 537 GQECVEBCRVLQGLPREYVNAHCLPCHPEQCP-QNGSVTCFGEADQCVCAHYKDPFF 595
Db 529 GKICIESCNLYDGEFREFENGSI-CVECDSDQCEKMEDEGLLTCHGCPDNCTCKSHFDPGN 588

Qy 596 CVARCPGVPDLSYMPIWKPDPDEGACQPCPINCTHSCVDLDDKGC-----PA 644
Db 589 CVEKCPDVLQANSF--IFKYADQDRCHPCHPNCTQCNGPTSHDCITYPMTGHTSLPQ 646

Qy 645 EQRASPLTSTVSVV-GILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPS 703
Db 647 HAR-TPL--TAAGVIGLFLVIMALTFAVTVRRRSIK-KRALRRFL-ETELVEPLTPS 701

Qy 704 GAMPNOAQMRIKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLENTSPKA 763
Db 702 GTPAQAQLRLIKETELRKVKVLSGAGFTVYKGIWVPEGETVXIPVAIKILNETTGPKA 761

Qy 764 NKEILDAYVMAGVGSYPVSRLLGICLTSTVLTQMLPYGCLLDHVRNGRGLSQDLL 823
Db 762 NVFEMDEALIMASVDHPHLVRLGLVCLSPITQLVTQMLPHGCLLEYVHEHKDNIQSOLL 821

Qy 824 NWCQIAGKMSYLEDVRLVHRDLAARNVLKNSPHNVKITDPLGLARLLDIDETEHADGCK 883
Db 822 NWCQIAGKMSYLEDVRLVHRDLAARNVLKNSPHNVKITDPLGLARLLDEGEKEYNADGCK 881

Qy 884 VPIKWMALLESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQ 943
Db 882 MPIKWMALCEHYRKFTHQSDVWSYGVTVWELMTFGKPYDGIIPREIPDLLEKGERLPQ 941

Qy 944 PPICTIDVYIMVKWIMIDSECRPRFRELVSFMRMARDPORFVVYIQNED-LGPASPLDS 1002
Db 942 PPICTIDVYIMVKWIMIDADRSRPFKELAAEFMRMARDPORIYIIOGDDRMKLPSPNDS 1001

Qy 1003 TFVBSLLEDDMDGLVDAEYLVPOQGFCCDP-----APCA 1039
Db 1002 KFFONLLEDEDDMDDAEYLVLP-QAFNIPPIYTSRTRIDSNRSEIGHSPPPAYTMS 1060

Qy 1040 GGMVHHRRHSSTRSGGDLTLGLEPSEEAAPRSP LAPSEGAGSDVFGDLGMGAAGLIQ 1099
Db 1061 GSQFVYQDGGPATQCG--MMPYTAITSTIPEAPVA--QGAETAEFDDSCNGTLRKPV 1115

Qy 1100 SLPHDPSPLORYSEDPTVPLPS-----ETDGYVAPLTCSPQPEYVQNPDRPQPPSP 1152
Db 1116 VPHQVEDSSTQRYSDPTVFAPERNRAELDEEYMTPMHDKPKQEYLNPE----- 1167

Qy 1153 REGPLAARPAGATLERAKTSLSPCKNGVKDVFAGFAGAVENPEYLTPOGGAAPQHPBP 1212
Db 1168 -ENPFVSR-----KNGDLQ-----ALDNPETHSASSG-----PPKA 1198

Qy 1213 -----FSPAFDNLNYDQDQPPPERGA--PFST 1236
Db 1199 EDEVNEPLYNTFTNALGNAEYMKNSLSVPEKAKAFDNDPDYNNHSLPRSTLQHPDY 1258

Qy 1237 FKGTPT-----AENPEYL 1249
Db 1259 LQESTYKFKQNGRIRPIVAENPEYL 1285
```

DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
 GN XMRK OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 NCBI_TaxID=8083;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=90015140; PubMed=2797166;
 RX Wittbrodt J., Adam D., Malitschek B., Maueier W., Raulf F.,
 RA Telling A., Robertson S.M., Scharf M.;
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RL inducing Tu locus in Xiphophorus."; Nature 341:415-421(1989).
 RN [2]
 RP REVISION TO 515.
 RA Scharf M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X16891; CAA34770.2; --
 CC PIR: S06142; S06142.
 DR HSP; P11362; IFGK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR PRINTS: PR00109; Recep_L_domain; 2.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 5.
 DR SMART: SM00220; S_TKC; 1.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE
 FT KINASE.
 FT DOMAIN 26 642
 FT TRANSMEM 643 665 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 666 1167 POTENTIAL.
 FT DOMAIN 710 977 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 716 724 ATP (BY SIMILARITY).
 FT BINDING 743 743 ATP (BY SIMILARITY).
 FT ACT_SITE 835 835 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 228 BY SIMILARITY.
 FT DISULFID 224 236 BY SIMILARITY.

FT DISULFID 237 245 BY SIMILARITY.
 FT DISULFID 241 253 BY SIMILARITY.
 FT DISULFID 256 265 BY SIMILARITY.
 FT DISULFID 269 296 BY SIMILARITY.
 FT DISULFID 300 311 BY SIMILARITY.
 FT DISULFID 315 330 BY SIMILARITY.
 FT DISULFID 333 337 BY SIMILARITY.
 FT DISULFID 504 513 BY SIMILARITY.
 FT DISULFID 508 521 BY SIMILARITY.
 FT DISULFID 524 533 BY SIMILARITY.
 FT DISULFID 537 553 BY SIMILARITY.
 FT DISULFID 556 569 BY SIMILARITY.
 FT DISULFID 560 577 BY SIMILARITY.
 FT DISULFID 593 615 BY SIMILARITY.
 FT DISULFID 618 626 BY SIMILARITY.
 FT DISULFID 622 634 BY SIMILARITY.
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 365 365 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 398 398 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;
 Query Match 38.9%; Score 2652.5; DB 1; Length 1167;
 Best Local Similarity 44.8%; Pred. No. 8.6e-134;
 Matches 569; Conservative 169; Mismatches 388; Indels 145; Gaps 31;
 QY 4 AALCRGLLIALLPPGAASL---QVCTGDMKRLRPASPETHLDMLRLHLYQGVQVQGN 59
 DB 8 AALLQ--LLLVLSISRCSTDPKVKCGTSNQMTM---LDNHYLKKKKYSGCNVVLN 62
 QY 60 LEITYLPTNASLFLQDIQEVQYVLIANHQQVPLQRLRIVRGTQLFEDNYALAVLDN 119
 DB 63 LEITYTQENQDLSFLOSIQEVGGYVLIANNEVSTPLVNLRLIRGONLYEGNFTLLVMSN 122
 QY 120 GDPNNNTTPVTGASPGSLRELQRLRSITELKGGVLIQRNPOLCYQDTILWKDIFKHNQL 179
 DB 123 YQK-NPSSP--DVYQVGLQLQLSLNLTLSGGVKVSHNPLLCNVETINMWDIVDKTSNP 179
 QY 180 ALTLIDNRSRACHPCSKGSRGWESSEDCQSLTRTVTCAGC-ARCKGPLPTDCCH 238
 DB 180 TNNLIPHAFERQCKCDHGVCSWAPGCHQCKFTKLLCAEQCNRRCKGPIIDCCNE 239
 QY 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
 DB 240 HCAGGCTGPRATDCLACRDFNDGTCCKTCTPPPKIYDIVSHQVVDNPNIKYTFGAACVKE 299
 QY 299 CPYNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMQYIKANSKFIG 358
 DB 300 CPSNVYVTE-GACVRSKCSAGMLEVD-ENGRSKCKPCDGVCPKVCDDGIGISL-SNTIAVN 356
 QY 359 ITEL-EFAGCKKIFGSLAFLPESFDGDP---ASNTAPOYIKANSKFIGITELTYLYISA 414
 DB 357 STNIRSFNCTKINGDIILNRNSFEGDPHYKIGTMDPEHL---WNLTTVKEITGYLVIMW 413
 QY 415 WPDLSPLDSVFQNLQVIRGILHNGAYS-LTLQGLGISWLGRLSLRSLRGLSLALIHNTH 473
 DB 414 WPNMTSLSVFQNLLEIRGRTTFSRGFSFVVQVRLHQLGLRSLRSLRGLSLALIHNTH 473
 QY 474 LCFVTVPHDQLFRPHQALLHTANRPEDECVGEGLACHQLCARGHGWGPGTQCVNCSQ 533
 DB 474 LRYANTINMRLFRSEDSQSIYDART-----ENQTCNNECEDSGCGPGTMCVSCSLH 526
 QY 534 FLRGQCEVEECRVLOGLPREYVNRHCLPCHPECPQNGSVTCFGEAQCACAHYKOP 593
 DB 527 VDRGRCVASCNLLQGEPREAQVDRGRCVQCHOECLVQTDLSLTCYGPGRANCSKSAHFQDG 586
 QY 594 PFCVARCPGSKVPLDSYMPIMKFPDEEGACQPCPINCTHSCVCDLDDKCGPAEQRASPLTS 653

```
Db 587 PQCIPRCPHGLGDGDLT-LWKYADKMGQCPCHONCTQCGSPGLSGRCD-IVSHSL 644
Qy 654 IVSAVGLLVVGLVGLIKRQOKIRKYTHRRLLQETELVEPTTPSGAMNQAMR 713
Db 645 AVGLVGLLITVIVALLIVLLRRRIK-RKRTIRCLLQEKELVEPLTPSGQAPNQLR 703
Qy 714 ILKETELRKVKVLGSGAGFTYKGIWIPDGNVAPIVAIKVLRNTSPKANKETLDEAYV 773
Db 704 ILKETEFKORVLGSGAGFTYKGLMNPDGNIRIPVAIKVLRNTSPKANKETLDEAYV 763
Qy 774 MAGVGSYVSRLLGICLTSTVQLVTLMPYGLCLLDHVRNKRGLSGDLLNCMQIAKGM 833
Db 764 MASVDHPCVRLGICLTSAVLVTLMPYGLCLLDYVRQHQERICGQWLLNCVQIAKGM 823
Qy 834 SYLEDVLRHDLAARNVLKSPHVKITDGLARLLDIDTEYHAGGKVPYIKWMALES 893
Db 824 NYLEERHLVRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGKVPYIKWMALES 883
Qy 894 ILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVTM 953
Db 884 ILQWYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKELIASVLENGERLPQPPICTIEVTM 943
Qy 954 IMVKCWMIDSECRPRFRELSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDD 1013
Db 944 IILKCMWIDPSSRRFRELVEGFESOMARDPSRYLVIOG---NLPSLSDRRLFSRLSSDD 1000
Qy 1014 MGLVDAAEYLVPOGFCFPPAPGACGGMVHRHSRSTRSGGGLDILGLEPSEEAAPS 1073
Db 1001 --DVVDADAYLLPYKRI-----NRQGS-----E 1021
Qy 1074 PLAPSEAGSDVDFDGLMGAAKGLQLPHTDPSPLQRYSDPTV-PLPSETDGVAPLT 1132
Db 1022 PCIPPTGH-----PVRENSITLNIISDPTQNALEKDLGDH----- 1056
Qy 1133 CSPOEYVNPQDVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTILSPGKNG 1179
Db 1057 -----EYVNPQGSTSSRLSDIYNPNVEDLTDGGMVPSLSQEAETNFSRPEYLTNQNS 1111
Qy 1180 VKVDVFAFGAVENPEYVLTPOGGAAPQHPPPAPSFADNLYYWDQPPERGAPPSTFKG 1239
Db 1112 L---PLVSSGSMDDPDY---QAG-----YQAAF-----LPOTGALTNGMF 1146
Qy 1240 TPTAENPEYLG 1250
Db 1147 LPAENLEYLG 1157

RESULT 9
ERB3 HUMAN
ID ERB3 HUMAN STANDARD; PRT; 1342 AA.
AC P21850;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90082324; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
```

```
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL tyrosine kinase.";
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAk.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -!- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M29366; AAA35790.1; -.
CC EMBL; M34309; AAA35979.1; -.
CC EMBL; S61953; AAB26935.1; -.
CC PIR; A36223; A36223.
CC HSP; P11362; 1FGK.
CC Genew; HGNC:3431; ERBB3.
CC MIM; 190151; -.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
CC PROSITE; PS00109; PROTEIN KINASE_TYR; FALSE_NEG.
CC PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Alternative splicing.
CC SIGNAL 1 19
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202
FT POTENTIAL.
FT RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT PROTEIN KINASE.
FT ATP (BY SIMILARITY).
FT ATP (BY SIMILARITY).
FT BY SIMILARITY.
FT BY SIMILARITY.
```

```

FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 141 183 EILSGGVYIEKNDKLCHMDTIDWRDIDVRDADAEIVKDNCR
SC -> GQFMVPSGLTPQQAQWYLLDDDDPRLLTLSASK
VPVTLAAV (IN SHORT ISOFORM).
MISSING (IN SHORT ISOFORM).
E -> G (IN REF. 2).
E -> G (IN REF. 2).
Query Match 34.9%; Score 2375.5; DB 1; Length 1342;
Best Local Similarity 40.0%; Pred. No. 5.3e-119;
Matches 525; Conservative 197; Mismatches 457; Indels 135; Gaps 34;
SEQUENCE 1342 AA; 148097 MW; 7201E7E66CA374BD CRC64;

```

```

QY 10 GLLALLPPGAA--STQVCTGTDMKLRPASPTHLDMLRHLHYQGVQVQGNLELTVLPT 67
DB 11 GLLFSLARGSEVGNQAVCPGLTNGLSVTGDAENQYQTLKYLCYCEVWGNLEIVLTGH 70
QY 68 NASLSPLDIOEVQGVYLIHQNQVQVPLQRURIVRGTLQFEDNYALAVLDNGDPLNNTT 127
DB 71 NADLSFQWIREVTGYVLVAMNEFSTLPLNLRVVRGTQVYDGKFAIFVM----LNYNT 125
QY 128 PVTGASPGGLRELRLSLTEILLKGGVLIQNPOLCYQDTILWKDIFHKNNQLALTIDTN 187
DB 126 ----NSSHALRQLRLTQLTEILSGGVYIEKNDKLCHMDTIDWRDIDVRDRD---AEIVVKD 178
QY 188 RSRACHPCSPMKGSRWCMSSEDCOSLRTRVCAGGC-ARCKGPLPTDCHEOCAGCTG 246
DB 179 NGRSCPPCHEVCKG-RCWGPGSEDCQTLTKTICAPQCNCHCFGNPNQCCHECAGCSCG 237
QY 247 PKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNYLST 306
DB 238 PQDTDCFACRHFNDSGACVPRCPQPLVYNKLTFLQLEPNPHTTKYQYGVGVCAVCPHFV-V 296
QY 307 DVGSCTLVCPLNQVETADGTQRCCKSPCARVCYGLGMQVIKANSKF--IGITELE- 363
DB 297 DQTSVVRACFPDKMEVD-NKGLMKCEPCGGLCPKACEGTG-----SGSRFQTVDSNIDG 350
QY 364 FAGCKKIFGSLAFPSFGDGPASNTAPOYIKANSKFIGITELTGVLVYISAMPDSIPDLS 423
DB 351 FVNTKILGNLDFLTGLNGDPWHKIPALDPKLVNVRTVREITGVLTNQSPPHMNF 410
QY 424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWGLRLSLRELGLSLJLHNNHLCFVHTVPW 482
DB 411 VFSNLTITGGRSLYNGFSLIMKLNLTSLGFRSLKEISAGRIYISANRQLCYHSLNW 470

```

```

QY 483 DQLFRNPHQALLHTA-NRPEDECVGEGLAQHQLCARHCWGPGTQCVNCQSFIRGQECV 541
DB 471 TKVLRGPTBERLDIKHNRPRRDCVABGVKCDPLCSSGGCMGPGPGQCLSCRNYSRGVCV 530
QY 542 EECRVLOGLPREVYNARHCLPCHPEQOPONGSVTCFGEADOCVACAHYHDPPECVACRP 601
DB 531 THCNFLNGEPREFAEAECSFCEPCOMEGTATCNGSGSDTCAQCAHFRDGPCHVSCSP 590
QY 602 SGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPABQRA----SPLTSTVSA 657
DB 591 HGVLG--AKGPIYKPDVQNECRPCHENCCTQCGKGPBLQDCLGQTLVLIGKTHLTMTV 648
QY 658 VVGILLVVLGVVFGILIKRQOKIR-KYTMRLLOTELVEPLTPSGAMPNQOMRILK 716
DB 649 IAG--LVWIFMLGGTFLYWRGRRIQNKRAMRRYLERGESIEPLDPS-ERANKVLARIFK 705
QY 717 ETELRRKVKVLSGCAFGTVVYKGIWIPGENVKIPIVAIKVLENTSPKANKELIDEAYVMAG 776
DB 706 ETELRLKVLGSGVFGTVHKGWIPGESIKIPVCIKVIDKSGRGOSFOAVTDHMLAIGS 765
QY 777 VGSPPVSRLLIGICTSTVOLTPQVGLDHHVRENRRGLSGODLLNMCWIAKMSYL 836
DB 766 LDHAHIVRLGLCPGSSLQVLYPLGSLLDHVRQHRGALGPQLLLNNGVQIAKGMYYL 825
QY 837 EDVRLVHRDLAARNVLVGNHVKITDFGLARLLDIDETEHADGGKVPKIKMALESILR 896
DB 826 EEHGMVHRNLAARNVLLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTPKIKMALESIFH 885
QY 897 RRFTHOSDVSQVTVWELMTFGCAKPDGIPAREIPDLLEKGERLPPOPTCTIDVYIMV 956
DB 886 GKYTHOSDVSQVTVWELMTFGAEPYAGRLAEVDPDLLEKGERLAQPOQCTIDVYIMV 945
QY 957 KCMWIDSECRPRELVSFSEFMRARDPQRFVVIQNEDLGPA---SPLDSTFYRSLLEDD 1013
DB 946 KCMWIDENIRTPFKELANEFTNARDPPRYLVIKRES-GGCIAPGEPHGLTNKKLEEVE 1004
QY 1014 MDGLVDAEYLVPOQGFCCPDPAAGAGMVHHRHSRSTSGGDDTLGLSP-SEEBAPR 1072
DB 1005 LEPELDLDLLEAEED-----NLATTLGSALSPLVGTILNRRGSQ 1045
QY 1073 SPLASEGAGSVDFDGLGKAAGLQSLPTHDP-PSPLQRYSEDPTVPLP-----SETD 1125
DB 1046 SLLSPSSGY-MPMNQNGLGESQESAVSGSSERCPRPVSLH-----PMRPGCLASESS 1098
QY 1126 GYVA-----PLTCSPOPE---YVNPQDVVRPQPPSPREGP----- 1156
DB 1099 GHVTGSEAELEQKVSRCRSRSPRPRCDSAYHSQRHSLLTPTVPLSPGPLEEEDVNG 1158
QY 1157 --LPAARPAGATLERAKTLSP-GKNGV-----KDVFAFGGAVENPEYLTPOGGAQP 1207
DB 1159 YVMPDTHLKGTPSSREGTLLSVGLSSVLGTEEBEED-----EEVEYMNRRRRHSP-P 1209
QY 1208 HPPAFSPAFDNLVYWD-----QDPPRGAPSTFTKGTPTAENPEYL 1249
DB 1210 HPPRPSLEELGYEYMDVGSLSASLGTSQSCPLHPVIMPTAGTTTDEDEYEM 1263

```

RESULT 10

ERB3_RAT

ID_ERB3_RAT

STANDARD;

PRT; 1339 AA.

AC Q62799; Q62955;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)

DE (c-erbB3).

GN ERB3.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

CC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
RT recombinant protein.";
RN Gene 165:279-284 (1995).
RN [2]
RP REVISIONS TO 85; 513 AND 565.
RA Hellyer N.J., Koland J.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuroligins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RN J. Neurosci. 17:1642-1659 (1997).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAk.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U29339; AAC28498.2; --
DR EMBL; U52530; AAC53050.1; --
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 644 662 POTENTIAL.
FT DOMAIN 663 1339 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 183 259 CVS-RICH.
FT DOMAIN 707 964 PROTEIN KINASE.
FT NP_BIND 713 721 ATP (BY SIMILARITY).
FT BINDING 740 740 ATP (BY SIMILARITY).
FT ACT_SITE 832 832 BY SIMILARITY.
FT DISULFID 186 194 BY SIMILARITY.
FT DISULFID 190 202 BY SIMILARITY.
FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.

FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1028 1028 L -> P (IN REF. 3).
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBDFDIE CRC64;

Query Match 33.8%; Score 2303.5; DB 1; Length 1339;
Best Local Similarity 40.3%; Pred. No. 3.6e-115;
Matches 518; Conservative 173; Mismatches 435; Indels 159; Gaps 35;

QY 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMKRLPASPETHDMLHLYQSCOVVQGN 59
DB 7 LQVLC----FLLSLARGSEMNSQAVCPGTGLNGLSVTGADNNOYQTLKLYEKEVVMGN 62
QY 60 LELYLTPTNASLFLQDIOEVQGYVLIHNNQVRQVPLQLRLIRIVRGTLQFEDNYALAVLDN 119
DB 63 LEIVLTGHNADLSFLQWIREVTGYVLVANNEFSVLPPLNLRVVRGTQVYDGKFAIFVM-- 120
QY 120 GDPLNNTTPVTGASPGGLRELQLRSITLKGVLQIRNPOLCYQDTILWKDIFHNQNL 179
DB 121 ---LNYNT---NSSHALRQLKFTQLTEILSGGVYIEKNDKLCMDTIDWRDVRVR-- 170
QY 180 ALTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTVACGCG-ARCKGRLPTDCCHE 238
DB 171 GAEIVVKNNGANCPPCHEVCKG--RCWGPDPDCQILTKTICAPQCNGRCGPNPQCCHD 229
QY 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTTFESMPNPEGRYTFGASCVTA 298
DB 230 ECAGGCGSPQDTCDFACRRFNDSGACVPRCPPELVYNKLTFFOLEPNHTKYQYGGVCVAS 289
QY 299 CPNYLSTDVGSCTLVCPHNOEVTADGTQRCCKSKPCARVCYGL--GMQYIKANSKF 356
DB 290 CPHNFV-VDQTFECVRACPPDKMEVD-KHGLKMCPCGGLCPKACEGTGSGSRYTQVDSN 347
QY 357 IGITELFAGCKKI FGSLLAFLESFDGDPDASNTAPOYIKANSKFIGITELTYLYISAWP 416
DB 348 ID----GFVNCTKILGNLDFLITGLNVDPWHKIPALDPEKLVNFRVREITGLYNTQSWP 403
QY 417 DSLPDLVSFONLQVIRGRILHNGAYS-LTLOGLIGISWLGRLSRLRELGSGLALIHNTLC 475
DB 404 PHMHFVSFNSLTTIGGRSLYNRGFSLLIMKNLVNLTSLGRSLKEISAGRVYISANQQLC 463
QY 476 FVHTVPWQDLFRNPQOALLHTA-NRPEDECVGEGLACHOLCARGHGWGPGPTQCVNCSQF 534
DB 464 YHSLNMTLLRGPSEERLDIKYDRPLGELAEKGVCDPLCSSSGGCGWPGPGQCLSCRNY 523
QY 535 LRGEQCEVERCVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDP 594
DB 524 SREGVCVTHCNFLQGEPRFVHEAQCFSCHEPCELMGEGTSTCNGSGSGSDACARCAHFRDGP 583
QY 595 FCVARCPGKVPDLSTYMPIWKFPDEEGACQPCPINCTHSC--VDLDKDGKPAQRASPLT 652


```
Db 584 HCVNSCPHGILG--AKGPIYKPAQNECRCHENCHENCTGCGNPELQDCLGQAEVLMSKPH 641
QY 653 SIYSAVVGILLVVLGVVGLIKRROOKIR-KYTMRLLOETELVEPLTPSGAMPNQAQ 711
Db 642 LVIAVTG--LAVILMILGSGFLYWRGRIQNKAMRYLGRGESIEPLDPS-EKANKVL 698
QY 712 MRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIIDEA 771
Db 699 ARIFKETELRKVLKVLGSGVFTVHKGIWIPGESIKIPVCIKVIEDKSGRQFOAVTDM 758
QY 772 YVMAGVSPYVSRLLGLICLTSTVOLVTOLMPYGCLLDHVRENRLGSDQLLNCMQIAK 831
Db 759 LAVGSLDHAHIVLGLICPGSSSLQVLYQLPLGSLLDHVHVKHRETLQPQLLLNNGVQIAK 818
QY 832 GMSYLEDLVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEXHADGKGVPIKKNAL 891
Db 819 GMYLLESHSWHRDLARNVWLKSPSQVQVADFGVADLLPPDDKQLLHSEAKTPIKKNAL 878
QY 892 ESILRRRFTHQSDVMSYGVTVWELMTGCAKPYDGIIPAREIPDLLEKGERLPQPPICTIDV 951
Db 879 ESIHFGKYTHQSDVMSYGVTVWELMTGCAEPYAGLRLAEIPDLLEKGERLAQPOICTIDV 938
QY 952 YMIWVKCWMIDSECRPFRELVSFSRMDPQRFVIONEDLGPASPLDSTFYRSLLED 1011
Db 939 YMWVKCWMIDENIRPTFKELANEFTRMARDPPRYLVIKRAS-GPGTP--PAAEPSVLTT 995
QY 1012 DDMGDLVDAEYLVPOQGFCCPDAPGAGGVHHRHSSSTRSCGGDLTLGLEPSEE--- 1068
Db 996 KEL-----QEALEPEL-----DLDLLEAESEGLA 1021
QY 1069 -----EAPRSPLAPSEG-----AGSDVDFDGLGMCAGKGLQSLPT 1103
Db 1022 TSLGSALLPTGTLTRPRGSQLSPSSGYMPNQSSIGEAELCLDSAVLGGREQSPRISL 1081
QY 1104 HDSPFLORYSEDTPVLPSETDGV-----APL-----TC-----SPQPE-----YVNPQ 1143
Db 1082 H-PIPRGR-----PASESEGHVTGSEALOEKVSVCRSRSPRPRGDSAYHSQR 1133
QY 1144 DVRPQPPSPREG-----LPAARPAGATLERAKTLP--GKGVV-----KDVF 1185
Db 1134 HSLTPTVPLSPGLEEDGNGYWPHTLARGASSREGTLSSVGLSVLGTDEED-- 1191
QY 1186 AFGAVENPEYLTPOGGAAPQPHPP 1210
Db 1192 -----EEYEYMKRRKRGSP-PRPP 1209

RESULT 11
EGFR DROME STANDARD; PRT; 1426 AA.
AC P0412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (Tortpodo protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RN [2]
RP REVISIONS.
```

```
RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A. CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; Tissue=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunko B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Mays A.D., Ibegwam C.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Li X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wagsarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.
RC STRAIN=Daekwanryeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S. III, Biloiseau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal
```

RT growth factor receptor.";
RL Nature 314:178-180(1985).
RN (8)
RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP ANALYSIS.
RX MEDLINE=92038942; PubMed=936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RT "Interallelic complementation among DER/flb alleles: implications for
RT the mechanism of signal transduction by receptor-tyrosine kinases.";
RN Genetics 129:191-201(1991).
RL (9)
RP REVIEW.
RX MEDLINE=97248481; PubMed=9094709;
RA Perrimon N., Perkins L.A.;
RT "There must be 50 ways to rule the signal: the case of the Drosophila
RT EGF receptor";
RL Cell 89:13-16(1997).
CC -!- WHICH BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF ANNIOSEROSA
CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC (r send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF052754; AAC08536.1; -;
DR EMBL; AF052753; AAC08536.1; JOINED.
DR EMBL; AF052754; AAC08535.1; -;
DR EMBL; AF052752; AAC08535.1; JOINED.
DR EMBL; K03054; AAA51462.1; -;
DR EMBL; K03417; AAA51460.1; -;
DR EMBL; K03416; AAA50965.1; -;
DR EMBL; K03418; AAA51461.1; -;
DR EMBL; AF109077; AAD26134.1; -;
DR EMBL; AF109078; AAD26132.1; -;
DR EMBL; AF109082; AAD26132.1; JOINED.
DR EMBL; AF109078; AAD26133.1; -;
DR EMBL; AF109084; AAD26133.1; JOINED.
DR EMBL; AF109079; AAD26130.1; -;
DR EMBL; AF109081; AAD26130.1; JOINED.
DR EMBL; AF109079; AAD26131.1; -;
DR EMBL; AF109083; AAD26131.1; JOINED.
DR EMBL; AF109080; AAD26135.1; -;
DR EMBL; AE003454; AAF46732.1; -;
DR EMBL; X02293; CAA26157.1; -;
DR EMBL; X78920; CAA55523.1; -;
DR EMBL; X78918; CAA55521.1; -;

DR EMBL; X78919; CAA55522.1; -;
DR PIR; A00640; GORPE.
DR HSSP; P11362; 1FGK.
DR Flybase; FBgn0003731; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW Developmental protein.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 869 889 POTENTIAL.
FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 938 1198 PROTEIN KINASE.
FT NP_BIND 944 952 ATP (BY SIMILARITY).
FT BINDING 971 971 ATP (BY SIMILARITY).
FT ACT_SITE 1063 1063 BY SIMILARITY.
FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
Query Match 28.2%; Score 1922; DB 1; Length 1426;
Best Local Similarity 32.4%; Pred. No. 7.3e-95;
Matches 464; Conservative 184; Mismatches 432; Indels 350; Gaps 43;
QY 24 QVCTGTMKRLPASPTHLDMLRHLYCGCQVVOGNELTYLTPT-NASLSFLQDIQEVQG 82
DB 100 KICIGTKSRLSVPSNKEHHYHNLDRDRTNCTYVDGNLKLTLWLPNENLDLSFLDNIREVTG 159
QY 83 YVLIAHNOVROVPLQRLIRIVRGTLF-----EDNYALAVLDNGDPLNNTTPTVTGASPGGL 137
DB 160 YILSHVDVKVFPKLIIRGRTLFLSLVSEVKYALFV-----TYSKM 203
QY 138 RELQRLSITEILKGGVLIQRPOLCYQDITLWKDFHKNNQLALTLDITNRSRACHPCSP 197
DB 204 YTLSEIPDLRDVLNGOVGFHNNYINLCHMRTIQMSEIVSNGTDAYVNYDFTAPERECPKCHE 263
QY 198 MCKGSRGWGSESDCQSLTRTVACAGGCA--RCKGPLPTDCCHCOCAAGCTGPKHSDCLAC 255
DB 264 SCTHG-CWEGGPKNCQKFSKLTCSPOCAGGRCYGPKPRECCHLFCAGGCTGTQKDCIAC 322
QY 256 LHFNHSGICELHCPALVTYNTDTFESMPNPGRYTTFGASCVTACPYNYLSTDVSGCTLVC 315
DB 323 KNFFDEAVSKECEPPMRKYNPTTYVLETNPGKYATGATCVKECP-GHLLRDNGACVRSC 381
QY 316 PLHNQEVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIGITEL-----EFAGCKK 369
DB 382 PQDKNDXGGE-----CVPCNGPCPKTC-----PGVTVLHAGNIDSRNCTV 422
QY 370 IFGSLAFLPESFDG--DPASN--TAPQVIKANSK----FIGITELTYLYISAWPDSLDP 421
DB 423 IDGNIRILDQTFSGFDQVYANYTNGPRYIPLDPERREVSTVKITGYLNIETGTHPQRN 482
QY 422 LSVFQNLQVIRILHNGAY--SLTLOGLGISWLGLRSLRGLSGLALTHHNTLCFVHTV 480
DB 483 LSYFRNLETHIGROLMESFAALAIVKSSLSYSEMRNLKQISSGSSVWIQHNRDLCYVSN 542
QY 481 PWQDLFRNPQALLHTANRPEDECVGSLACHQLCARGHCWGPPTQCVNCQSLRGQEC 540
DB 543 RWPAIQKEPEQKVVNENLRADLCCKNGTICSDQCNEDGCMWAGTDCGLTCKNFNFNGTC 602
QY 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFSGPEADQCACAHYKDPFCVVARC 600

```
Db 603 IADCGYISAYK--FDNRCKICHPECR-----TCNGAGADHCOECVHVVRDQGHCVCSEC 654
Qy 601 P-----SGVK-----PDL 608
Db 655 PKNYNDRGVCRECHATCDGCTGPKDITIGAGCTTCNLAIINNDATVRCRLKDKDCPD- 713
Qy 609 SYMPIWKF--PDEGACOP-----CPI-----NCTH-----632
Db 714 QY--FWEVHFQEGSLKPLAGRAVCRKCHPLCELCCTNYGHEQVCSKCTHYKRQCET 771
Qy 633 -----SC-----VDLDKG-----641
Db 772 ECPADHYTDEORECFQRHPECNGCTGPGADDDCKSCRNFKLFANETGTYVNVSTMFNCTS 831
Qy 642 -CPAEQR-----ASPLTS-----IVSAVVGILLVVVLGVVGI 673
Db 832 KCPLMRHVNYQYTAIGPYCAASPRSSKITANLDVNMFIITGAVLVPTICILCV--T 889
Qy 674 LIKROQKIRKYT--MRLLQETELVEPLTPSGAMPNOAMRILKETELRKVKVLGSGAF 731
Db 890 YICQKQAKKETVMTWALSGCESEPLRFSNTGANKLCKRIVKDAELRGVGLMGAF 949
Qy 732 GTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVSPYVSRLLGICLT 791
Db 950 GRVYKGVWVPSGENVKIPVAIKELKSTGAESSEEFLEAYIMASEEHVNLKLLAVCMS 1009
Qy 792 STVOLVTOLMPYGCILLDHVRENRLGSGDILLNWCQIAKMSYLEDVRLVHRDLAARNV 851
Db 1010 SQMMLITQLMPLGCLLDVVRNRRDKIGSKALLINWSTQIAKMSYLEEKRLVHRDLAARNV 1069
Qy 852 LVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWMALESILRRTHQSDVWSYGV 911
Db 1070 LVQTPSLVKITDFGLAKLLSSDSNEYKAAGGKMPKWLALCEIRNRVFTSKSDVWAFGV 1129
Qy 912 VWELMTGAKPYDGIAPAREIPDLLEKGBRLPOPICTIDVYVMIMVKWMIDSECRPRFRE 971
Db 1130 IWELLTFQORPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCHWLDAAEPTFKQ 1189
Qy 972 LVSEFSRWARDPQRFVWQNEGLG--PASPLDSTFYRSLEDD-----DMGDLVDAEYLV 1026
Db 1190 LTTVFAEPDPRGYLAIPGDKFTLPA-----YTSQDEKDLIRKLAPTDTGSEAIK 1242
Qy 1027 QQGFCEPDPAACAGMVHHRSSSTRSGGDLTLGLEPSEEAPE-----RSPLAPSEG 1080
Db 1243 PDDYLQPKAAFGPS-----HRTDCT-----DEMPKLNRYCKDPSNKNSS 1281
Qy 1081 AGSDVFDG---DLGMAAGKLQSLPHTDPSPLQRYSEDPVLPSETDGYVAPLTCSPQP 1137
Db 1282 TGDDESDSSAREVGVGNLR-----LDLPVDEDDYLMPTCQPGP 1319
Qy 1138 EYVNOPVVRPQPPSPREGPLPAARPAATLRAKTLSPGKNGVVKDVFAGGAVENPEYL 1197
Db 1320 NNNNMN-----NPNQNNMAAVGAAGY-----DLIGVPVSDVNDPEYL 1358
Qy 1198 ----TPQGAAPQH-----PPAFSP-AFONLYYND 1224
Db 1359 LNAOTLVGESPIPTQTIGIPVMGPGTMEVKVPMGSEPTSSDHESYND 1408

RESULT 12
ERBB ALV
ID ERBB ALV
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DT Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB
OS Avian leukosis virus.
OC Viruses; Retroviridae; Retrovirus.
OX NCBI_TaxID=11864;
RN (1)
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC
CC -1- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFLV.
DR HSSP; P11362; 1FCG.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR TRANSFERASE; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.7%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 4.3e-86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;

Qy 587 CAHYKDPFPFCVAPCPGVKPDLSYMPIWKFPDEECACOPCPINCTHSCVDLDDKGCPEAQ 646
Db 3 CAHFIDGPHCVKACPAVGLGENDTL-VWKYADANAVCQLCHNCNTRCGKPGLEGCP--- 58
Qy 647 RASPLTSTVSAVV-GILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGA 705
Db 59 NGSKTPSIAAGVVGGLLCLVVGVLGIGLYLRRR-HIVRKRTLRLQLRELVEPLTPSGE 117
Qy 706 MPNOAMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANK 765
Db 118 APNQAHRLKETEFKVKVILGSGAFGVYKGLWIPEGEKVKIPVAIKELREATSPKANK 177
Qy 766 EILDEAYVMAGVSPYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSGDOLLNW 825
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVQLVTQMLPYGCLLDYIREHKDNGISQVLLNW 237
Qy 826 CQIAKMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVP 885
Db 238 CVQIAKGMVLEERLVRDLAARNVVKTPQHVKITDFGLAKLLGADEKEYHAEGKVP 297
Qy 886 IKWMALESILRRTHQSDVWSYGVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLPQP 945
```

or send an email to license@sib-ch).

EMBL: M20386; AAA48760.1; InterPro: IPR000494; EGFR_L domain. InterPro: IPR000719; Euk_Pkinase. InterPro: IPR002174; Furin-like. Pfam: PF00757; Furin-like; 1. Pfam: PF01030; Recep_L_domain; 2. SMART: SMO0261; FU; 4. PROSITE: PS00107; PROTEIN KINASE ATP; PARTIAL. PROSITE: PS00109; PROTEIN KINASE TYR; PARTIAL. PROSITE: PS00011; PROTEIN KINASE DOM; PARTIAL. Transmembrane: Glycoprotein; Receptor; Signal. Tyrosine-protein kinase; ATP-binding; Phosphorylation. SIGNAL 1 30 CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR. DOMAIN 31 654 EXTRACELLULAR (POTENTIAL). TRANSMEM 655 667 POTENTIAL. DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL). FT DISULFID 197 206 BY SIMILARITY. FT DISULFID 201 214 BY SIMILARITY. FT DISULFID 222 230 BY SIMILARITY. FT DISULFID 226 238 BY SIMILARITY. FT DISULFID 239 247 BY SIMILARITY. FT DISULFID 243 255 BY SIMILARITY. FT DISULFID 258 267 BY SIMILARITY. FT DISULFID 271 298 BY SIMILARITY. FT DISULFID 302 314 BY SIMILARITY. FT DISULFID 318 333 BY SIMILARITY. FT DISULFID 336 340 BY SIMILARITY. FT DISULFID 513 522 BY SIMILARITY. FT DISULFID 517 530 BY SIMILARITY. FT DISULFID 533 542 BY SIMILARITY. FT DISULFID 546 562 BY SIMILARITY. FT DISULFID 565 581 BY SIMILARITY. FT DISULFID 569 589 BY SIMILARITY. FT DISULFID 592 601 BY SIMILARITY. FT DISULFID 605 627 BY SIMILARITY. FT DISULFID 630 638 BY SIMILARITY. FT DISULFID 634 646 BY SIMILARITY. FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 633 633 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL). FT NON_TER 703 703 N-LINKED (GLCNAC. . .) (POTENTIAL). SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;

Query Match 22.8%; Score 1553; DB 1; Length 703; Best Local Similarity 43.9%; Pred. No. 1.4e-75; Matches 311; Conservative 109; Mismatches 258; Indels 30; Gaps 14;

8 RWGLLLALLPPGAA-----STOVCTGDMKLRLPASPETHLDMLRLHYOGCOVVOGNLE 61
13 RGAALVLLLLGVALCSAVEKKVCGQTNKLTQLGHVEDHFTSLQRYNCEVVLNLE 72
62 LTYLPTNASLSFLQDIOEVQGYVLIANOVQVPLQRLIVRGQTQFPEDNYALAVLNGD 121
73 ITVEHNRDLTLFKTIQEVAGYVLIANVMDVPLENLQIRGNVLVDNSFALAVLSNYH 132
122 PLNNTTPVTGASPGGLRELQRLSLTELKGGVLIQORNPOLCYODTILWKDIFHKNNQAL 181
133 -MUKTQ-----GLRELPMKRLSELNGVKISNNPKLCNMDTVLWNDIIDTSRK-PL 182
182 TLID-TNRSRACHPCSPMCKGRCWGESSEDCQSLRTVTCAGCA-RCKGPLPTDCHEQ 239

Search completed: July 22, 2003, 08:46:09
Job time : 20.2304 secs

183 TVLDFASNISSCPKHPNCTEDHCWAGEONCOTLTKVICAQCSGRCKRGKVPSDCCNQ 242
240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRITTFGASCYTAC 299
243 CAAGCTGPRESDCLACRKRFRDDATCKTCPLVLYNPTTYQMDVNPBGYSFCATCVR 302
300 PNYLSTDVGSCTLVCPILHNOEVTAEQDCEKSKPCARVCYGLGMOYIKANSKFIGI 359
303 PHNYVVTVDHSCVRSCTNTDTTEV-EENGVRKCKKCDGLCSKVCNIGIGELGKILS-INA 360
360 TELE-FAGCKKIFGSLAPLPESFDGDPASNTAPQYIKANSKFIGITELTGLYVISA 418
361 TNIDSFNCKTKINGDVSILPVAFLGDAFTKLPDPKKLDVFTVKEISGELLQAWPDN 420
419 LPDLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSLRELSSGLALHNNTHLCFVH 478
421 ATDLAYAFENLEIRGRTKQHGYSLAVNLIKISLGLSLKEISDGDIAIMKNKNCVAD 480
479 TVPDOLFNRPHQALLHTANRPEDECVGEGGLACHQICARGHCGPGTQCVCNCSOFLRGQ 538
481 TMNRSLFATQSKTKIIONRNKNDCTADRHVCDPLCSDVGCWGPFGPFHCFRCRFSRQK 540
539 ECVESCRVLQGLPREYVVARHCLPCHPECPONG---SVTCFGEADQCACAHYKDPDF 595
541 ECVKCNILQGEPRERFERSKCLPCHSECLVQNTSTAYNTTCSGPGPDHCKMCAHFDGPH 600
596 CVARCPGKVPDLSVMPITWKFDEEGACQPCPINCTHSCVDLDDKGCAPABORASPLTSIV 655
601 CVKACPAGVLGENDTL-VMKYADANAVCQLCHPNCTRGCKGPGLEGCP---NGSKTPSIA 656
656 SAVV-GILLVVVLGVVFGILIKRQOKIRKYVMRRLLOETELVEPLTP 702
657 AGVVGGLLCVVGVLGIGLYLRRR-HIVKRTLRLLQERELVEPLTP 703

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-325-339-12
Perfect score: 6814
Sequence: 1 MELAAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6734	98.8	1255	21	Human heregulin 2
2	6734	98.8	1255	22	Human tyrosine kin
3	6734	98.8	1255	22	HER2 transgene pla
4	6734	98.8	1255	23	Human HER2 (ErbB2)
5	6728	98.7	1255	17	Human HER2/neu protein.
6	6728	98.7	1255	20	Human HER-2/neu on
7	6728	98.7	1255	21	Amino acid sequenc
8	6728	98.7	1255	21	Human HER-2/neu pr
9	6728	98.7	1255	22	Human HER-2/neu pr
10	6728	98.7	1255	22	HER2/neu amino aci

11	6728	98.7	1255	23	AAE24067	Human Her-2 protei
12	6728	98.7	1255	23	AAE20479	Human Her-2/neu pr
13	6728	98.7	1255	23	AAE20479	Human Her-2/neu pr
14	6728	98.7	1255	23	AAU71143	Human Her-2/neu po
15	6685	98.1	1433	14	AAE39568	Sequence of c-erbB
16	6564	96.3	1223	23	AAU98923	Human breast cance
17	6411	94.1	1200	21	AAE21208	Human HER-2/neu pr
18	5941.5	87.2	1256	21	AAE21199	Rat Her-2/neu prot
19	5941.5	87.2	1256	21	AAE21199	Rat Her-2/neu onco
20	5914.5	86.8	1256	21	AAE21206	Mouse Her-2/neu pr
21	5914.5	86.8	1256	21	AAE21206	Mouse Her-2/neu pr
22	5914.5	86.8	1256	23	AAE21206	Amino acid sequenc
23	4814	70.6	919	21	AAE21203	Mouse Her-2/neu on
24	4814	70.6	919	21	AAE21203	Human HER-2/neu fu
25	4064.5	59.6	920	23	AAE21152	Mouse Her-2/neu ex
26	4064.5	59.6	926	23	AAE21152	Mouse Her-2/neu ex
27	3698	54.3	712	21	AAE21204	Human HER-2/neu fu
28	3698	54.3	712	21	AAE21204	Human HER-2/neu fu
29	3552	52.1	782	18	AAW19764	Her2-GM-CSF immuno
30	3550	52.1	653	21	AAE21200	Extracellular HER-
31	3550	52.1	653	23	AAE21200	Human Her-2/neu on
32	3512	51.5	645	22	AAE60408	Human ErbB2 oncopr
33	3512	51.5	645	22	AAE61593	Human ErbB2 extrac
34	3447	50.6	951	21	AAE44993	DC8cFV-erbB2EC fu
35	3344	49.1	624	11	AAE08222	Extracellular port
36	3146	46.2	1210	21	AAE19259	Amino acid sequenc
37	3146	46.2	1210	21	AAE19259	Human EGF receptor
38	3146	46.2	1210	23	AAE23019	Human Her-1 protei
39	3146	46.2	1210	23	AAE50768	Human epidermal gr
40	3144	46.1	1210	22	AAE68420	Amino acid sequenc
41	3105	45.6	1210	23	AAE19768	Human epidermal gr
42	3084	45.3	583	23	AAE20483	Human protein for
43	3084	45.3	587	23	AAE20481	Human protein for
44	3083	45.2	589	23	AAE20484	Human protein for
45	3083	45.2	600	23	AAE20482	Human protein for

ALIGNMENTS

RESULT 1
AAE20482
ID AAY92620 standard; Protein; 1255 AA.
XX AAY92620;
AC AAY92620;
DT 10-AUG-2000 (first entry)
XX Human heregulin 2 (Her2).
DE Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain /label= N-terminal
FT /note= "mature polypeptide"
FT Region /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Domain /label= insertion region
FT /note= "suitable for foreign epitope insertion"

Qy 781 YVSLGICLTSTVQLTQVLTQMPYCGCLLDHVRNRRGLSGQDLNKNWQIAKMSYLEVDV 840
Db 781 YVSLGICLTSTVQLTQVLTQMPYCGCLLDHVRNRRGLSGQDLNKNWQIAKMSYLEVDV 840
Qy 841 LVHRDLARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMMALESILRRFT 900
Db 841 LVHRDLARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKWM 960
Qy 961 IDSECRPRFRELUSEFMRMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVA 1020
Db 961 IDSECRPRFRELUSEFMRMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVA 1020
Qy 1021 EYLVPQOGFCPPDPAPGAGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPQOGFCPPDPAPGAGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGLGMGAAGLQSLPTHDPSPLOQRYSEDPTVLPSETDGVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMGAAGLQSLPTHDPSPLOQRYSEDPTVLPSETDGVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQSPREGPLPAARPAGATLERAKTLSPGKNGVVKOVFAFGAVENPEYLTPO 1200
Db 1141 NOPDVRPQSPREGPLPAARPAGATLERAKTLSPGKNGVVKOVFAFGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPPPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPPPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 2

AAE12130

ID AAE12130 standard; Protein; 1255 AA.

XX AC AAE12130;

XX DT 18-DEC-2001 (first entry)

XX DE Human tyrosine kinase-type receptor, HER-2.

XX KW Therapeutic compound; major histocompatibility complex; vaccine;

XX KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;

XX KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;

XX KW antigen presenting cell; human; tyrosine kinase-type receptor.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Region 774..782

XX FT /note= "Antigenic epitope"

XX PN WO200168677-A2.

XX PD 20-SEP-2001.

XX PP 16-MAR-2001; 2001WO-US40328.

XX PR 16-MAR-2000; 2000US-0527487.

XX XX (GENZ) GENZYME CORP.

XX XX Nicolette CA;

XX DR WPI; 2001-616284/71.

XX DR N-PSDB; AAD19731.

XX PT Novel synthetic therapeutic compound for inducing immune response and

XX PT for use in adoptive immunotherapy, has enhanced binding to major

XX PT histocompatibility molecules and enhanced immunoregulatory properties

PT

XX

PS Claim 4; Page 63-67; 69pp; English.

CC The invention relates to synthetic therapeutic compounds (antigenic peptides) with enhanced binding to major histocompatibility complex (MHC) molecules and enhanced immunoregulatory properties relative to their natural counterparts. Compounds of the invention are useful for inducing an immune response in a subject and for use in adoptive immunotherapy. They are useful as components of anti-cancer vaccines and to expand immune effector cells that are specific for cancers characterised by expression of the breast cancer antigen, HER-2.

CC Polynucleotides that encode peptides of the invention are useful as hybridisation probes and as primers for the detection of genes of gene transcripts that are expressed in antigen presenting cells (APCs), to confirm transduction of polynucleotides into host cells. The present sequence is human tyrosine kinase-type receptor, HER-2. Compounds of the invention are designed based on the HER-2 antigenic peptide (774-782).

XX

SQ Sequence 1255 AA;

Query Match 98.8%; Score 6734; DB 22; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLRPASPTHDLMLRHLVYOGCVVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLRPASPTHDLMLRHLVYOGCVVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVQGVYLI AHNOVRQVPLQRLIRVGRQTQFEDNYALVDNG 120

Db 61 ELTYLPTNASLSFLQDIOEVQGVYLI AHNOVRQVPLQRLIRVGRQTQFEDNYALVDNG 120

Qy 121 DPLNNTTPTVTCGSPGGLRELQSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNOLA 180

Db 121 DPLNNTTPTVTCGSPGGLRELQSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNOLA 180

Qy 181 LTLDITNRSRACHPCSPMKCGSRGWSSSDCQSLTRTVACAGGCARCKGPLTDCCHQOC 240

Db 181 LTLDITNRSRACHPCSPMKCGSRGWSSSDCQSLTRTVACAGGCARCKGPLTDCCHQOC 240

Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVSGCTLVCPPLHNOEVTAEQTCRCKSKPCARVCYGLGHQYIKANSKFIOIT 360

Db 301 YNYLSTDVSGCTLVCPPLHNOEVTAEQTCRCKSKPCARVCYGLGHQYIKANSKFIOIT 360

Qy 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEETITGYLYISAWPDSLP 420

Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEETITGYLYISAWPDSLP 420

Qy 421 DLSVFQNLQVIRGRIILHNGAYSILTQGLGISWGLSLRELGLSLALIHNNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRIILHNGAYSILTQGLGISWGLSLRELGLSLALIHNNTHLCFVHTV 480

Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGPTQCVNCSQFLRGOEC 540

Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGPTQCVNCSQFLRGOEC 540

Qy 541 VEECRVLQGLPREYVVARHCLPCHPECOPONGSVTCFGEADOCVCAHYKDPFCVVARC 600

Db 541 VEECRVLQGLPREYVVARHCLPCHPECOPONGSVTCFGEADOCVCAHYKDPFCVVARC 600

Qy 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTISAVVG 660

Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTISAVVG 660

Qy 661 ILLVVLGVVFGVGLIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720

Db 661 ILLVVLGVVFGVGLIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720


```

OY 721 RKVKVLSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVGP 780
DB 721 RKVKVLSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVGP 780
OY 781 YVSRLLGICLTSTVOLVTQMLPYPGCLLDHVRENRLGSLQDLLNMCQIAKMSYLEVDV 840
DB 781 YVSRLLGICLTSTVOLVTQMLPYPGCLLDHVRENRLGSLQDLLNMCQIAKMSYLEVDV 840
OY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRFT 900
OY 901 HQSDVWSYGVTVWELMTFCAPYDGIPIAREIPDLLEKGERLPQPPICITDVTMIMVKWM 960
DB 901 HQSDVWSYGVTVWELMTFCAPYDGIPIAREIPDLLEKGERLPQPPICITDVTMIMVKWM 960
OY 961 IDSECRPRELVSFSESMARDPQRFVVIQNEIDLGPASPLDSTFVRSLLDDDDMDGLVDA 1020
DB 961 IDSECRPRELVSFSESMARDPQRFVVIQNEIDLGPASPLDSTFVRSLLDDDDMDGLVDA 1020
OY 1021 EYLVPOQGFPCDPAPGAGGVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
DB 1021 EYLVPOQGFPCDPAPGAGGVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
OY 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPQLRYSEDTVPLPSETDGYVAPLTCSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPQLRYSEDTVPLPSETDGYVAPLTCSPQPEYV 1140
OY 1141 NQPDVVRPQPPSPREGPLPAARPAATLERAKTLPSPKNGVVKDVFAGAVENPEYLTPO 1200
DB 1141 NQPDVVRPQPPSPREGPLPAARPAATLERAKTLPSPKNGVVKDVFAGAVENPEYLTPO 1200
OY 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
DB 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
AC AAB60167;
DT 03-APR-2001 (first entry)
DE HER2 transgene plasmid construct encoded protein.
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
OS Homo sapiens.
OS Synthetic.
XX WO200100244-A2.
PN 04-JAN-2001.
XX 23-JUN-2000; 2000WO-US17229.
PD 25-JUN-1999; 99US-0141316.
PF 16-MAR-2000; 2000US-0189844.
PR (GETH ) GENENTECH INC.
XX Erickson S, Schwall R;
XX WPI; 2001-061962/07.
DR N-PSDB; AAF24297.
XX Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX

```

```

PS Example 3; Fig 4; 92pp; English.
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 98.8%; Score 6734; DB 22; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
OY 1 MELAALCRWGLLLALLPPGAASQTCTGTDMLRLPASPEHLDMLRLHYQCGVQVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASQTCTGTDMLRLPASPEHLDMLRLHYQCGVQVQGNL 60
OY 61 ELTYLPTNASLFLQDIOEVQGYVLIHNOVROVPLQRLRIVRGTFQLEDNYALAVLDNG 120
DB 61 ELTYLPTNASLFLQDIOEVQGYVLIHNOVROVPLQRLRIVRGTFQLEDNYALAVLDNG 120
OY 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRPOLQCYQDTILWKDIFHKNNQLA 180
DB 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRPOLQCYQDTILWKDIFHKNNQLA 180
OY 181 LTLIDTNRSRACHPCSPCKGSRWGSESSBQSLTRTVAGGCARCKGPLPTDCHEQC 240
DB 181 LTLIDTNRSRACHPCSPCKGSRWGSESSBQSLTRTVAGGCARCKGPLPTDCHEQC 240
OY 241 AAGCTGPKHSGDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
DB 241 AAGCTGPKHSGDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
OY 301 YNYLSTVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLMGVYKANSKFIGIT 360
DB 301 YNYLSTVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLMGVYKANSKFIGIT 360
OY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEBITGYLISAMPDSL 420
DB 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEBITGYLISAMPDSL 420
OY 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLISWLGRLSRLGSLGIALIHNHNLCEVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLISWLGRLSRLGSLGIALIHNHNLCEVHTV 480
OY 481 PWDQLFRPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC 540
DB 481 PWDQLFRPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC 540
OY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPPEFCVARC 600
DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPPEFCVARC 600
OY 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCFAQRASPLTSIVSAVVG 660
DB 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCFAQRASPLTSIVSAVVG 660
OY 661 ILLVVVLGVWFGLIKRRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAORILKETEL 720
DB 661 ILLVVVLGVWFGLIKRRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAORILKETEL 720
OY 721 RKVKVLSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVGP 780
DB 721 RKVKVLSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVGP 780
OY 781 YVSRLLGICLTSTVOLVTQMLPYPGCLLDHVRENRLGSLQDLLNMCQIAKMSYLEVDV 840
DB 781 YVSRLLGICLTSTVOLVTQMLPYPGCLLDHVRENRLGSLQDLLNMCQIAKMSYLEVDV 840
OY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRFT 900

```

Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVKPIKMALESILRRPT 900
 Qy 901 HQSDVMSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPPCTIDVTYIMVKCM 960
 Db 901 HQSDVMSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPPCTIDVTYIMVKCM 960
 Qy 961 IDSECRPRFRELVSFSESWARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDMDGLVDA 1020
 Db 961 IDSECRPRFRELVSFSESWARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDMDGLVDA 1020
 Qy 1021 EYLVPOQGFCDPAPGAGGMVHRHRSSTFRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 Db 1021 EYLVPOQGFCDPAPGAGGMVHRHRSSTFRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 Qy 1081 AGSDVDFDGLNGAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Db 1081 AGSDVDFDGLNGAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Qy 1141 NOPDVRPQPPSPREGPLPAARAGATLERAKTLSPGKGVKDVAFGAGAVENPEYLTPO 1200
 Db 1141 NOPDVRPQPPSPREGPLPAARAGATLERAKTLSPGKGVKDVAFGAGAVENPEYLTPO 1200
 Qy 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
 Db 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 4

ID AAU74545
 XX AAU74545 standard; Protein; 1255 AA.
 AC AAU74545;
 DT 23-APR-2002 (first entry)
 DE Human HER2 (ErbB2) polypeptide.
 KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glioma disorder; astrocytoma disorder; hypothalamic disorder;
 KW glandular disorder; macrophage disorder; epithelial disorder;
 KW stomal disorder; blastocoele disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.
 XX Homo sapiens.
 OS
 XX
 PN US2002001587-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 16-MAR-2001; 2001US-0811123.
 XX
 PR 16-MAR-2000; 2000US-189844P.
 PR 05-OCT-2000; 2000US-238327P.
 XX
 PA (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIW/) SLIWKOWSKI M.
 XX
 PI Erickson S, Schwall R, Sliwowski M;
 XX
 DR WPI; 2002-163686/21.
 DR N-PSDB; ABK14058.
 XX
 PT Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor. ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
 XX
 XX Example 3; Fig 7; 93pp; English.
 XX
 CC The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytic, hypothalamic, glandular, macrophage,
 CC epithelial, stromal, blastocoele, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.

XX SQ Sequence 1255 AA;

Query Match 98.8%; Score 6734; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPFGAASSTQVCTGDMKRLPASPETHLMLRHLYQGCQVQGNL 60
 Db 1 MELAALCRWGLLLALLPFGAASSTQVCTGDMKRLPASPETHLMLRHLYQGCQVQGNL 60
 Qy 61 ELTYLPTNASLFLQDIOEVQGYVLIHNNQVQVPLQRLIRVGTQQLFEDNVALAVLDNG 120
 Db 61 ELTYLPTNASLFLQDIOEVQGYVLIHNNQVQVPLQRLIRVGTQQLFEDNVALAVLDNG 120
 Qy 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGVLIQRPOLCYQDTILWKDIFHKNOLA 180
 Db 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGVLIQRPOLCYQDTILWKDIFHKNOLA 180
 Qy 181 LTLDTNRSRACHPCSPCKSGSRGSESDCQSLTRTVACGCGARCKGLPTDCCHEQC 240
 Db 181 LTLDTNRSRACHPCSPCKSGSRGSESDCQSLTRTVACGCGARCKGLPTDCCHEQC 240
 Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTAC 300
 Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTAC 300
 Qy 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTORCEKSCPCARVCYGLGCMQYIKANSKEFIGIT 360
 Db 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTORCEKSCPCARVCYGLGCMQYIKANSKEFIGIT 360
 Qy 361 ELEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEQLQVLFETLGTLYISAMPDSL 420
 Db 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEQLQVLFETLGTLYISAMPDSL 420
 Qy 421 DLSVFNQLQVIRGRILHNGAYSLTLQGLISWLGRLSRLGSLGLALIHNNTHLCFVHTV 480
 Db 421 DLSVFNQLQVIRGRILHNGAYSLTLQGLISWLGRLSRLGSLGLALIHNNTHLCFVHTV 480
 Qy 481 PWDQLFRPHOALLHTANRPEDECYGEGLACHQLCARGHCWPGPTQCVNCSQFLRGQEC 540
 Db 481 PWDQLFRPHOALLHTANRPEDECYGEGLACHQLCARGHCWPGPTQCVNCSQFLRGQEC 540
 Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
 Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
 Qy 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCPCAPQASPLTSIVSAVVG 660
 Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCPCAPQASPLTSIVSAVVG 660
 Qy 661 ILLVVVLGVVFGIILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPQNAQRILKETEL 720
 Db 661 ILLVVVLGVVFGIILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPQNAQRILKETEL 720
 Qy 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAIVMAGVGS 780
 Db 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAIVMAGVGS 780
 Qy 781 YVSRLLGLICTSTVQLVTQLMPYGLCLDHPVRENRRGLSQDILLNMCQIAKMSYLEBVR 840

Db 781 YVSRLLGICLTSTVQLVTQLMFYGCLLDHVRNRCGLSGQDLLNMCQIAKMSYLEYDR 840
Qy 841 LVHRLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALSILRRRT 900
Db 841 LVHRLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALSILRRRT 900
Qy 901 HOSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMIWVKWM 960
Db 901 HOSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMIWVKWM 960
Qy 961 IDSECRPRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDMDGLVDA 1020
Db 961 IDSECRPRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDMDGLVDA 1020
Qy 1021 EBYLVPOQFFCPDPAAGAGGVHHRSSSTRSGGDLTLGLEPSEEAAPSPAPSEG 1080
Db 1021 EBYLVPOQFFCPDPAAGAGGVHHRSSSTRSGGDLTLGLEPSEEAAPSPAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGKLSLTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEV 1140
Db 1081 AGSDVFDGDLGMAAGKLSLTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEV 1140
Qy 1141 NOPDVRPOPSPREGPLPAARPAAGATLERAKTSLSPGKNGVWVDVPAFGAVENPEYLTQ 1200
Db 1141 NOPDVRPOPSPREGPLPAARPAAGATLERAKTSLSPGKNGVWVDVPAFGAVENPEYLTQ 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 5

AAW01111
ID AAW01111 standard; Protein; 1255 AA.
XX
AC AAW01111;
XX
DT 01-JAN-1997 (first entry)
XX
DE HER-2/neu protein.
XX
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 676..1255
FT /label= Intracellular domain
FT /note= "Claimed domain, useful for immunisation"
XX
PN W09630514-A1.
XX
PD 03-OCT-1996.
XX
PF 28-MAR-1996; 96WO-US01689.
XX
PR 31-MAR-1995; 95US-0414417.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Cheever MA, Disis ML;
XX
DR WPI; 1996-455361/45.
DR N-PSDB; AAT40739.
XX
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
XX associated
XX
PS Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transfected host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 98.7%; Score 6728; DB 17; Length 1255;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPFGAASQVCTGTDMLRLPASPTHLDMLRLHYQGCVVQGNL 60
Db 1 MELAALCRWGLLALLPFGAASQVCTGTDMLRLPASPTHLDMLRLHYQGCVVQGNL 60
Qy 61 ELYLPTNASLSFLQDIQEVQGYVLIHNVQVQVPLQRLIRVGTQLPEDNYALVDNG 120
Db 61 ELYLPTNASLSFLQDIQEVQGYVLIHNVQVQVPLQRLIRVGTQLPEDNYALVDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCVQDTILWKDIFHKKNOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCVQDTILWKDIFHKKNOLA 180
Qy 181 L7LIDTNRSRACHPCSPCKGSRGWGESSEDCQSLTRTVACGACARCKGPLPTDCCHEQC 240
Db 181 L7LIDTNRSRACHPCSPCKGSRGWGESSEDCQSLTRTVACGACARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRVTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRVTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLCVPLHNOEVTABDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
Db 301 YNYLSTDVGSCTLCVPLHNOEVTABDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFLESFDGDPASNTAPLQPEQLQVLFETLEEITGLYISAWPDSLP 420
Db 361 IQEFAGCKKIFGSLAFLESFDGDPASNTAPLQPEQLQVLFETLEEITGLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRLSRLSGLALIHNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRLSRLSGLALIHNTHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHOLCARGHCMGPGPTQCVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHOLCARGHCMGPGPTQCVCNCSQFLRGQEC 540
Qy 541 VECERVLOGLPREYVYNAHRLCHPCEQPQNGSVTCFGEADQCACAHYKDPFFCVARC 600
Db 541 VECERVLOGLPREYVYNAHRLCHPCEQPQNGSVTCFGEADQCACAHYKDPFFCVARC 600
Qy 601 PSGVKPDLISYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
Db 601 PSGVKPDLISYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
Qy 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLSGAFGTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYMAGVGSF 780
Db 721 RKVKVLSGAFGTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYMAGVGSF 780
Qy 781 YVSRLLGICLTSTVQLVTQLMFYGCLLDHVRNRCGLSGQDLLNMCQIAKMSYLEYDR 840
Db 781 YVSRLLGICLTSTVQLVTQLMFYGCLLDHVRNRCGLSGQDLLNMCQIAKMSYLEYDR 840

QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRFT 900
 DB |||||
 QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRFT 900
 DB |||||
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIFDLLEKGERLPQPICTIDVYMIWVKCM 960
 DB |||||
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIFDLLEKGERLPQPICTIDVYMIWVKCM 960
 DB |||||
 QY 961 IDSECRPRFRELVSFSRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDMDGLVDA 1020
 DB |||||
 QY 961 IDSECRPRFRELVSFSRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDMDGLVDA 1020
 DB |||||
 QY 1021 EBYLVPOQGFPCDPAPGAGGVVHRHRSSTRSGGDLTLGLEPSESEAPSLAPSEG 1080
 DB |||||
 QY 1021 EBYLVPOQGFPCDPAPGAGGVVHRHRSSTRSGGDLTLGLEPSESEAPSLAPSEG 1080
 DB |||||
 QY 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEV 1140
 DB |||||
 QY 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEV 1140
 DB |||||
 QY 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTSLSPGKGVVVDVFAFGAVENPEYLTQP 1200
 DB |||||
 QY 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTSLSPGKGVVVDVFAFGAVENPEYLTQP 1200
 DB |||||
 QY 1201 GGAAPQPPHPPAFSPAFDNLVYWDODPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
 DB |||||
 QY 1201 GGAAPQPPHPPAFSPAFDNLVYWDODPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
 DB |||||

RESULT 6

ID AAW92406
 ID AAW92406 standard; Protein; 1255 AA.
 AC AAW92406;
 XX
 XX
 DT 21-APR-1999 (first entry)
 XX
 DE Human HER-2/neu oncogene protein.
 XX
 KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 KW malignancy; treatment; tumour.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 676..1255
 FT /note= "region which elicits immune response"
 XX
 XX US5869445-A.
 XX
 XX 09-FEB-1999.
 XX
 XX 01-APR-1996; 96US-0625101.
 XX
 XX 01-APR-1996; 96US-0625101.
 PR 17-MAR-1993; 93US-0033644.
 PR 12-AUG-1993; 93US-0106112.
 PR 31-MAR-1995; 95US-0414417.
 XX
 XX (UNIW) UNIV WASHINGTON.
 XX
 XX Cheever MA, Disis ML;
 XX
 XX WPI; 1999-152835/13.
 DR N-PSDB; AAX01912.
 XX

Use of HER-2/neu polypeptides - for eliciting an immune response to
 an HER-2/neu associated malignancy, particularly for treating or
 preventing tumours

Claim 3; Column 31-38; 26pp; English.

This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or reoccurrence.
 XX

Sequence 1255 AA;

Query Match 98.7%; Score 6728; DB 20; Length 1255;
 Best Local Similarity 98.6%; Pred No. 0;
 Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MELAALCRWGLLLALLPFGAASTQVCTGDMKRLPASPETHLDMRLHYQGVVQGNL 60
 DB 1 MELAALCRWGLLLALLPFGAASTQVCTGDMKRLPASPETHLDMRLHYQGVVQGNL 60
 QY 61 ELTYLPTNASISFLQDIOEVQGVYLIANQVROVPLQRLIRVGTQLFEDNALAVLDNG 120
 DB 61 ELTYLPTNASISFLQDIOEVQGVYLIANQVROVPLQRLIRVGTQLFEDNALAVLDNG 120
 QY 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQRPOLCVQDTILWKDIFHKNNOLA 180
 DB 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQRPOLCVQDTILWKDIFHKNNOLA 180
 QY 181 LTLIDTNRSRACHPCSPCKSGRCGSESDCQSLTRTVACGACRCKGLPTDCCHEQC 240
 DB 181 LTLIDTNRSRACHPCSPCKSGRCGSESDCQSLTRTVACGACRCKGLPTDCCHEQC 240
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
 QY 301 YNYLSTDVGSCTLCVPLHNQEVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIGIT 360
 DB 301 YNYLSTDVGSCTLCVPLHNQEVTAEDGTORCEKSKPCARVCYGLGMHLEHREVRAVTSAN 360
 QY 361 ELEFAGCKKIFGSLAFPESEFDGDPASNTAPLQPEQLQVLFETLEITGYLISAWPDSL 420
 DB 361 IQEFAGCKKIFGSLAFPESEFDGDPASNTAPLQPEQLQVLFETLEITGYLISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLETLQGLGISWGLRLSRLGSLGIALIHNHNLFCVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSLETLQGLGISWGLRLSRLGSLGIALIHNHNLFCVHTV 480
 QY 481 PWDOLFRNPHOALLHTANRPEDECVGEGLAHCGLCARGHCWGPCTQVCNCSQFLRGQEC 540
 DB 481 PWDOLFRNPHOALLHTANRPEDECVGEGLAHCGLCARGHCWGPCTQVCNCSQFLRGQEC 540
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
 QY 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIYSAVVG 660
 DB 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIYSAVVG 660
 QY 661 ILLVVVLGVVFGIILIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 DB 661 ILLVVVLGVVFGIILIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEAYVMAGVSP 780
 DB 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEAYVMAGVSP 780
 QY 781 YVSLLLGICLTSTVQLVTOLMPYGCLLDHVRENRLGSLQDLNLCWQIAKMSYLEVDV 840
 DB 781 YVSLLLGICLTSTVQLVTOLMPYGCLLDHVRENRLGSLQDLNLCWQIAKMSYLEVDV 840
 QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRFT 900
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRFT 900

QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYMIWVKWM 960
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYMIWVKWM 960
QY 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLLEDDMGDLVDA 1020
DB 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLLEDDMGDLVDA 1020
QY 1021 BEYLVPQCGFFCPDPAPGAGGMVHHRSSSTRSGGDLTTLGLEPSEEEAPRSLAPSEG 1080
DB 1021 BEYLVPQCGFFCPDPAPGAGGMVHHRSSSTRSGGDLTTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAKGLQSLPSPQLQRYSEDTVPPLPSETDGYVAPLTCSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAAKGLQSLPSPQLQRYSEDTVPPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NOPDVRPQPPSPREGPLPAARAGATLERAKTSLSPGKNGVVDVFAFGAVENPEYLTPO 1200
DB 1141 NOPDVRPQPPSPREGPLPAARAGATLERAKTSLSPGKNGVVDVFAFGAVENPEYLTPO 1200
QY 1201 CGAAPQPPPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
DB 1201 CGAAPQPPPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 7

AAB21198
ID AAB21198 standard; protein; 1255 AA.
XX AAB21198;
AC AAB21198;
DT 12-JAN-2001 (first entry)
XX Human HER-2/neu protein.
DE Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.
XX Homo sapiens.
XX WO200044899-A1.
XX 03-AUG-2000.
XX 28-JAN-2000; 2000WO-US02164.
XX 29-JAN-1999; 99US-0117976.
XX (CORI-) CORIXA CORP.
XX (SMIK) SMITHKLINE BEECHAM.
PI Cheever MA, Gheysen D;
DR WPI; 2000-505976/45.
DR N-P5DB; AAA89736.
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX Claim 52; Fig 7; 128pp; English.

CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
XX
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6728; DB 21; Length 1255;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPFGAASSTQCTGDMKRLRASPETHLDMRLHLYQGCQVQGNL 60
DB 1 MELAALCRWGLLLALLPFGAASSTQCTGDMKRLRASPETHLDMRLHLYQGCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLRI VRGTQOLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLRI VRGTQOLFEDNYALAVLDNG 120
QY 121 DPLNNTTVPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
DB 121 DPLNNTTVPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCOSLTRTVCAGGCARCKGPLTDCCHEOC 240
DB 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCOSLTRTVCAGGCARCKGPLTDCCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360
DB 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKCSKPCARVCYGLGMEHLREVRAVTSAN 360
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPQLQVFEITLBEITGYLISAMPDSL 420
DB 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPQLQVFEITLBEITGYLISAMPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSELGSLALIHNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSELGSLALIHNTHLCFVHTV 480
QY 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQF LRGQEC 540
DB 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQF LRGQEC 540
QY 541 VEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFGEADOCVACAHYKPPFCVARC 600
DB 541 VEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFGEADOCVACAHYKPPFCVARC 600
QY 601 PSGVKPDLSPYMPIWKFDPDEEGACQPCINCTHSCVDLDDKGCAPAQASPLTSIVSAVVG 660
DB 601 PSGVKPDLSPYMPIWKFDPDEEGACQPCINCTHSCVDLDDKGCAPAQASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGIILIKRQOKIRKYTMRLRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
DB 661 ILLVVVLGVVFGIILIKRQOKIRKYTMRLRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
DB 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
QY 781 YVSRLLGLICTSTVQLVTQLMPYGCGLDHVRENRLGSLQDLLNMCWIAKGMSTYLEDVR 840
DB 781 YVSRLLGLICTSTVQLVTQLMPYGCGLDHVRENRLGSLQDLLNMCWIAKGMSTYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP IKWMALESILRRRPT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP IKWMALESILRRRPT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYMIWVKWM 960
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYMIWVKWM 960
QY 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLLEDDMGDLVDA 1020
DB 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLLEDDMGDLVDA 1020

Db 961 IDSECRPRFRELSEFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLEDDMDGLVDA 1020
Qy 1021 EYLVPOQGFCCPPAPAGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPOQGFCCPPAPAGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGWAAGLQSLPTTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGWAAGLQSLPTTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLPSPKNGVVDVFAFGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLPSPKNGVVDVFAFGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAFTFGTPTAENPEYLGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAFTFGTPTAENPEYLGLDVPV 1255

RESULT 8
AAY84780
ID AAY84780 standard; Protein; 1255 AA.
XX
AC AAY84780;
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX
KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.
XX
OS Homo sapiens.
XX
PN WO200020579-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00912.
XX
PR 02-OCT-1998; 98US-0165192.
XX
PA (UYMC-) UNIV MCMASTER.
XX
PI Muller WJ, Siegel PM;
XX
DR WPI; 2000-303768/26.
DR N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPLICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6728; DB 21; Length 1255;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
Qy 1 MELAALCRWGILLALLPFGAASSTOCTGCTDMKRLRASPETHLDMRLHYQCCVQVQNL 60
Db 1 MELAALCRWGILLALLPFGAASSTOCTGCTDMKRLRASPETHLDMRLHYQCCVQVQNL 60
Qy 61 ELTYLPTNASLFLQDIOEVQYVLIHNOVROVPLQRLRIRVGTQLPEDNTALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIOEVQYVLIHNOVROVPLQRLRIRVGTQLPEDNTALAVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWSESSEBQSLTRTVACGACARCKGLPTDCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSRGWSESSEBQSLTRTVACGACARCKGLPTDCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFIPESPDGPASNTAPLOPELOVFPETLEBITGYLISAWPDSL 420
Db 361 IQEFAGCKKIFGSLAFIPESPDGPASNTAPLOPELOVFPETLEBITGYLISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYS LTLQIGISLWGLRSRLSGSLALHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYS LTLQIGISLWGLRSRLSGSLALHNNHLCFVHTV 480
Qy 481 PWDQLFRNPHOALLHTANRPEDECVGEGLAHQLCARGHCGPGTQCVCNSQFIRGQEC 540
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGLAHQLCARGHCGPGTQCVCNSQFIRGQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPPECVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPPECVARC 600
Qy 601 PSQVGPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKCPAQASPLTSIISAVVG 660
Db 601 PSQVGPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKCPAQASPLTSIISAVVG 660
Qy 661 ILLVVVLGVVFGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVSP 780
Db 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVSP 780
Qy 781 YVSRLLGICLTSTVOLVTLMPYGLLDHVRNRRGLSGODLANKMCIAGMSYLEDV 840
Db 781 YVSRLLGICLTSTVOLVTLMPYGLLDHVRNRRGLSGODLANKMCIAGMSYLEDV 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDYVMIMVKCM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDYVMIMVKCM 960
Qy 961 IDSECRPRFRELSEFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLEDDMDGLVDA 1020
Db 961 IDSECRPRFRELSEFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLEDDMDGLVDA 1020
Qy 1021 EYLVPOQGFCCPPAPAGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080

Db 1141 NQDVRQPPSPREGPLPAARPAGATLERPKTLSPKNGVVDVPAFCGAVENPEYLTPQ 1200
AAG88267
Qy 1201 GGAAPQHPHPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
|||||
Db 1201 GGAAPQHPHPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
|||||
RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
XX AAG88267;
XX
XX 11-SEP-2001 (first entry)
XX
XX HER2/neu amino acid sequence.
XX
XX Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
XX immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
XX tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
XX Homo sapiens.
XX
XX WO200141787-A1.
XX
XX 14-JUN-2001.
XX
XX 11-DEC-2000; 2000WO-US33591.
XX
XX 10-DEC-1999; 99US-0458299.
XX
XX (EPIM-) EPIMUNE INC.
XX
XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
XX Keogh E;
XX
XX WPI; 2001-374995/39.
XX
XX
XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
XX cellular immune responses for the prevention and treatment of cancer -
XX
XX Disclosure; Page 15; 199pp; English.
XX
XX The present invention describes isolated prepared HER2/neu epitopes (I).
XX Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
XX culture in vitro and binds to a complex of an epitope (I), bound to a
XX human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
XX and a second epitope and the peptide is less than 50 contiguous amino
XX acids that have 100% identity with a native peptide sequence of HER2/neu;
XX (3) a vaccine composition (III) comprising (II) and a pharmaceutical
XX excipient; (4) an isolated nucleic acid encoding a peptide comprising
XX (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
XX and immunostimulant activities, and can be used in vaccines. (I), (II)
XX and (III) are useful for inducing cellular immune responses for the
XX prevention and treatment of cancer. (I) and (II) are useful for
XX monitoring or evaluating an immune response to a tumour-associated
XX antigen when incubated with a T lymphocyte sample form a patient and
XX detecting the presence of bound T lymphocyte to (I) or (II). Epitope
XX based vaccines mean that immunosuppressive epitopes that may be present
XX in whole antigens may be avoided. Selected epitopes may be combined to
XX enhance immunogenicity. The possible pathological side effects caused by
XX infectious agents or whole protein antigen is eliminated. The vaccine
XX provides the ability to direct and focus an immune response to multiple
XX selected antigens from the same pathogen. Epitope-based anti-tumour
XX vaccines provides the opportunity to combine epitopes derived from
XX multiple tumour-associated molecules addressing the problem of tumour-
XX tumour variability and reducing the likelihood of tumour escape due to
XX antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
XX the exemplification of the present invention.
XX
XX Sequence 1255 AA;
Query Match 98.7%; Score 6728; DB 22; Length 1255;

Best Local Similarity 98.6%; Pred. No. 0;
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPFGAASQVCTGTDMLRLPASPEHLDMLRHLQCCQVVOGRL 60
|||||
Db 1 MELAALCRWGLLALLPFGAASQVCTGTDMLRLPASPEHLDMLRHLQCCQVVOGRL 60
|||||
Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNNQVRQVPLQRLRIVRGTLFEDNYALAVLNG 120
|||||
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNNQVRQVPLQRLRIVRGTLFEDNYALAVLNG 120
|||||
Qy 121 DPLNNTTPVTCASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
|||||
Db 121 DPLNNTTPVTCASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
|||||
Qy 181 LTLDTNRSRACHPCSPCKGRCWGESSEDCQSLTRTVACAGGCARCKGPLPTDCHEOC 240
|||||
Db 181 LTLDTNRSRACHPCSPCKGRCWGESSEDCQSLTRTVACAGGCARCKGPLPTDCHEOC 240
|||||
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
|||||
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
|||||
Qy 301 YNYLSTDVGSCTLVCPHLNQVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
|||||
Db 301 YNYLSTDVGSCTLVCPHLNQVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360
|||||
Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFPETLEITCYLYISAWPDSL 420
|||||
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFPETLEITCYLYISAWPDSL 420
|||||
Qy 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLIGISWGLRLSRLSGSLALIHNNTHLCFVHTV 480
|||||
Db 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLIGISWGLRLSRLSGSLALIHNNTHLCFVHTV 480
|||||
Qy 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQICARGHCNGPGTQCVCNSQFURGQEC 540
|||||
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQICARGHCNGPGTQCVCNSQFURGQEC 540
|||||
-Y 541 VEECRVLQGLPREVYNARHCLPCHPEQONGSVTCGPEADQCVACAHYKDPPECVARC 600
|||||
Db 541 VEECRVLQGLPREVYNARHCLPCHPEQONGSVTCGPEADQCVACAHYKDPPECVARC 600
|||||
Qy 601 PSGVKPDLSPYPIWKFPDEGACQPCPINCTHSCVDLDDKGCPEAQASPLTISVAVVG 660
|||||
Db 601 PSGVKPDLSPYPIWKFPDEGACQPCPINCTHSCVDLDDKGCPEAQASPLTISVAVVG 660
|||||
Qy 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
|||||
Db 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
|||||
Qy 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEAYVMAGVSP 780
|||||
Db 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEAYVMAGVSP 780
|||||
Qy 781 YVSRLLGICLTSTVQLVTLMPYCCLLDHVRENRGLGSDLLNWCMIAGMSYLEVDV 840
|||||
Db 781 YVSRLLGICLTSTVQLVTLMPYCCLLDHVRENRGLGSDLLNWCMIAGMSYLEVDV 840
|||||
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKWMALLESILRRFT 900
|||||
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKWMALLESILRRFT 900
|||||
Qy 901 HQSDVNSYGVTVWELMTFGAKPDVGIIPAREIPDLLEKGERLPQPICTIDYVMIMVKWM 960
|||||
Db 901 HQSDVNSYGVTVWELMTFGAKPDVGIIPAREIPDLLEKGERLPQPICTIDYVMIMVKWM 960
|||||
Qy 961 IDSECRPRFRELVESEFSRMDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDGLVDA 1020
|||||
Db 961 IDSECRPRFRELVESEFSRMDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDGLVDA 1020
|||||
Qy 1021 EYLVPOQGFCCPDPAAGAGMWHRRSSRSGGGDLTLGLEPSEEEAPRSLAPSE 1080
|||||

Db 1021 EYLVPOGPFCDPAPGAGGMMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGLQSLPTDPSPLQRYSEDPTVPLPSETDGYVAPITCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPTDPSPLQRYSEDPTVPLPSETDGYVAPITCSPQPEYV 1140
Qy 1141 NOPDVRPOPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Db 1141 NOPDVRPOPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 11

AAE24067
ID AAE24067 standard; Protein; 1255 AA.
XX
AC AAE24067;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human Her-2 protein.
XX
KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
XX
OS Homo sapiens.
XX
PN WO200222636-A1.
XX
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowseert LM;
XX
DR WPI; 2002-471192/50.
XX
DR N-PSDB; AAD38904.
XX
PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX
PS Example 13; Page 95-107; 116pp; English.
XX

CC The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6728; DB 23; Length 1255;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPPGAASSTQVCTGDMKLRLPASPETHLMDLRHLHYGCGVQVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASSTQVCTGDMKLRLPASPETHLMDLRHLHYGCGVQVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIQEVGYVLIHANOVRQVPLQRLRIVRGTQLPEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQDIQEVGYVLIHANOVRQVPLQRLRIVRGTQLPEDNYALAVLNG 120
Qy 121 DPLNNTTPVTGASPGGLREQLRLSLEILKGGVLIQRNPQLCYQDTILKWKDIFHKNNOLA 180
Db 121 DPLNNTTPVTGASPGGLREQLRLSLEILKGGVLIQRNPQLCYQDTILKWKDIFHKNNOLA 180
Qy 181 LTLDTNRSRACHPCSPMCKGSRGWSSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLDTNRSRACHPCSPMCKGSRGWSSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLHNOEVTADGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPLHNOEVTADGTQRCCKSKPCARVCYGLGMEHLREAVRTSAN 360
Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFPETLEETITCYLYISAMPDSL 420
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFPETLEETITCYLYISAMPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRLSRLSGLLHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRLSRLSGLLHNNTHLCFVHTV 480
Qy 481 PWDQLFRNPHOALLHTANRPEDECVGEGLACHQLCARGHCWPGPTQCVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGLACHQLCARGHCWPGPTQCVCNCSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPCVARC 600
Qy 601 PSGVKPDLSPYMPIWKFPDEGACQPCINCTHSCVDLDDKGCBAQORASPLTSISAVVG 660
Db 601 PSGVKPDLSPYMPIWKFPDEGACQPCINCTHSCVDLDDKGCBAQORASPLTSISAVVG 660
Qy 661 ILLVVVLGVVFGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLGSGARFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIIDEAYVMAGVSP 780
Db 721 RKVKVLGSGARFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIIDEAYVMAGVSP 780
Qy 781 YVSRLLGICLTSTVQVLTQMLMPYGLLDHVRNRLGSGDQLLNCMQLAKGMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQVLTQMLMPYGLLDHVRNRLGSGDQLLNCMQLAKGMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCM 960
Qy 961 IDSECRPRFRELVSFSESRMARDPQFVYVQNEIDLGASPLDSTFVRSILLEDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFSESRMARDPQFVYVQNEIDLGASPLDSTFVRSILLEDDMDGLVDA 1020
Qy 1021 EYLVPOQGPFCDPAPGAGGMMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPOQGPFCDPAPGAGGMMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGLQSLPTDPSPLQRYSEDPTVPLPSETDGYVAPITCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPTDPSPLQRYSEDPTVPLPSETDGYVAPITCSPQPEYV 1140
Qy 1141 NOPDVRPOPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200

Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVVKDVFAGGAVENPEYLTTPQ 1200
Qy 1201 GGAAPQHPPPAFSPAFDNLYWQDPPPERGAPSTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPPPAFSPAFDNLYWQDPPPERGAPSTFKGTPTAENPEYLGIDVPV 1255
RESULT 12
AAE20479 ID AAE20479 standard; Protein; 1255 AA.
XX AC AAE20479;
XX DT 01-JUL-2002 (first entry)
XX DE Human Her-2/neu protein.
XX KW Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
XX KM human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Region 1021..1030
XX FT /note= "Naturally processed HLA-B44-restricted epitope"
XX PN W0200214503-A2.
XX PD 21-FEB-2002.
XX PF 14-AUG-2001; 2001WO-US41733.
XX PR 14-AUG-2000; 2000US-225152P.
XX PR 28-SEP-2000; 2000US-236428P.
XX PR 21-FEB-2001; 2001US-270520P.
XX PA (CORI-) CORIXA CORP.
XX PI Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
XX PI McNeill PD, Vedvick TS;
XX DR WPI; 2002-280758/32.
XX DR N-PSDB; AAD32743.
XX PT Novel isolated Her-2/Neu polypeptide composition useful for therapy,
XX PS prevention and diagnosis of cancer, preferably breast cancer -
XX PS Disclosure; Page 114-117; 129pp; English.
XX CC The invention relates to an isolated Her-2/Neu polypeptide composition
XX CC effective for eliciting an immune response. The invention is useful for
XX CC eliciting an immune response in a patient, where the patient is human
XX CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
XX CC The composition is useful for the therapy and diagnosis of cancer,
XX CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
XX CC and other compositions for the diagnosis, prevention and treatment of
XX CC human malignancies, for stimulating and/or expanding T cells specific for
XX CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
XX CC patient. The invention is useful for stimulating a T cell response in a
XX CC human patient, as probe or primer for nucleic acid hybridisation, to
XX CC selectively form duplex molecules with complementary stretches of the
XX CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
XX CC length gene from a suitable library, and to direct expression of a
XX CC polypeptide in appropriate host cells. The composition is useful in
XX CC prophylactic or therapeutic applications and for the treatment of cancer,
XX CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
XX CC associated malignancies. The invention is useful in gene therapy. The
XX CC present sequence is human Her-2/neu protein.
XX SQ Sequence 1255 AA;
Query Match 98.7%; Score 6728; DB 23; Length 1255;

Best Local Similarity 98.6%; Pred. No. 0;
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLPASPTHLDMLRHLGCGVVOGNL 60
Db 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLPASPTHLDMLRHLGCGVVOGNL 60
Qy 61 ELYLPTNASLSFLQDIOEVGYVLIHNOVRQVPLQRLRIVRGTLQEDNYALAVLNG 120
Db 61 ELYLPTNASLSFLQDIOEVGYVLIHNOVRQVPLQRLRIVRGTLQEDNYALAVLNG 120
Qy 121 DPLNNTTPTGASPGGLREQLRLSLTEILKGGVLIORNPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTTPTGASPGGLREQLRLSLTEILKGGVLIORNPOLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMCKGSRGWGSSDQSLTRTVCAAGGACRCKPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMCKGSRGWGSSDQSLTRTVCAAGGACRCKPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCITLVCPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCITLVCPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 ELEPAGCKKIFGSLAFLPESPDGDPASNTAPLOPEQLQVETLEITCYLYISAWPDSL 420
Db 361 IQEPAGCKKIFGSLAFLPESPDGDPASNTAPLOPEQLQVETLEITCYLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRIHNGAYSILTLOGLISWLGRLSRLGSLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRIHNGAYSILTLOGLISWLGRLSRLGSLALIHNNTHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAACHQICARGHCHGPGPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAACHQICARGHCHGPGPTQCVNCSQFLRGQEC 540
Qy 541 VECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
Db 541 VECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPAEQASPLTSIVAVVG 660
Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPAEQASPLTSIVAVVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETILDEAYVWAGVSP 780
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETILDEAYVWAGVSP 780
Qy 781 YVSRLLGICLITSTVQLVTQMLPYGCLLDHVRNRRGLSGDLDLNCWQIAGMSYLEVDVR 840
Db 781 YVSRLLGICLITSTVQLVTQMLPYGCLLDHVRNRRGLSGDLDLNCWQIAGMSYLEVDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPIKWMALLESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPIKWMALLESILRRFT 900
Qy 901 HQSDVMSYGVTVWELTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYMIWVCWM 960
Db 901 HQSDVMSYGVTVWELTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYMIWVCWM 960
Qy 961 IDSECRPRFRELVSFESRWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFESRWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
Qy 1021 EBYLVPQOQFFCDDPAPGAGGMVHRRHSSTSGGGDLTLGLEPSEEEAPRSLAPSEG 1080

Db 1021 EYLVPQQFFCPDPAAGAGGMVHRRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGWAAGLQSLPTDHPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGWAAGLQSLPTDHPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVFAGGAVENPEYLTTPQ 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERPKTSLSPGKNGVVKDVFAGGAVENPEYLTTPQ 1200
Qy 1201 GGAAPQHPHPPAFSPAFNLVYDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFNLVYDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 13

AAM51143

ID AAM51143 standard; Protein; 1255 AA.

XX AC AAM51143;

XX DT 17-JUN-2002 (first entry)

XX DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

XX KW tyrosine kinase; receptor; c-erbB2; gene therapy.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Domain 1..653

FT FT /note= "extracellular domain"

FT Domain 676..1255

FT FT /note= "intracellular domain"

FT Domain 990..1255

FT FT /note= "phosphorylation domain"

XX WO200212341-A2.

XX PN 14-FEB-2002.

XX PD 03-AUG-2001; 2001WO-US24283.

XX PF 03-AUG-2000; 2000US-0632507.

XX PR (CORI-) CORIXA CORP.

XX XX (SMWK) SMITHKLINE BEECHAM BIOLOGICALS.

XX FI Cheever MA, Gheysen D;

XX XX WPI; 2002-241743/29.

XX DR N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT or enhancing an immune response to the protein, has Her-2/neu
PT extracellular domain fused to Her-2/neu intracellular or
PT phosphorylation domain -
XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
CC c-erbB2), an oncogenic self-protein and target for anti-cancer
CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
CC in a variety of cancers, including breast, ovarian, colon, lung and
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
CC family of receptor-like glycoproteins. It comprises an extracellular
CC domain with homology to the epidermal growth factor receptor
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC intracellular domain that also shows homology to EGFR. Its
CC overexpression correlates with a poor prognosis in breast and
CC ovarian cancers. The invention provides Her-2/neu fusion
CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its DeltaPD fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 98.7%; Score 6728; DB 23; Length 1255;

Best Local Similarity 98.6%; Pred. No. 0;

Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASSTQCTGTDMLRLPASPETHLDMRLHYQCCVVOGNL 60

Db 1 MELAALCRWGLLLALLPPGAASSTQCTGTDMLRLPASPETHLDMRLHYQCCVVOGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEQVGYVLIAHNQVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIOEQVGYVLIAHNQVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120

Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

Qy 181 LTLIDTNRSRACHPCSPMKCGSSESCOSLTRVCAGGCARCKPLPTDCCHEQC 240

Db 181 LTLIDTNRSRACHPCSPMKCGSSESCOSLTRVCAGGCARCKPLPTDCCHEQC 240

Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360

Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEITGYLISAMPDSL 420

Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEITGYLISAMPDSL 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSRLSGSLALIHNNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSRLSGSLALIHNNTHLCFVHTV 480

Qy 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQCHGCPGPTQCVCNSQFIRGQBC 540

Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQCHGCPGPTQCVCNSQFIRGQBC 540

Qy 541 VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFPEADQCACAHYKDPPEVCVARC 600

Db 541 VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFPEADQCACAHYKDPPEVCVARC 600

Qy 601 PSGVKPDLSPYMPIWKPFDEEGACOPCPINCTHSCVDLDDKGPAPORASPLTISIVAVVG 660

Db 601 PSGVKPDLSPYMPIWKPFDEEGACOPCPINCTHSCVDLDDKGPAPORASPLTISIVAVVG 660

Qy 661 ILLVVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Db 661 ILLVVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVWAGVGP 780

Db 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVWAGVGP 780

QY 781 YVSRLLGICLTSTVOLVTLMPYGCCLLDHVRENRLGSGDQLLNWCMQIAKMSYLEVDV 840
DB 781 YVSRLLGICLTSTVOLVTLMPYGCCLLDHVRENRLGSGDQLLNWCMQIAKMSYLEVDV 840
QY 841 LVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRFT 900
DB 841 LVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRFT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
QY 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
DB 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
QY 1021 EYLVLVPOQFFCPDPAFCAGGMVHRHRSSTRSGGGDLTLGLEPSEEEAPRSLAPSEG 1080
DB 1021 EYLVLVPOQFFCPDPAFCAGGMVHRHRSSTRSGGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGKQLSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAAGKQLSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
DB 1141 NOPDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
DB 1201 GGAAPQHPHPPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 14

AAU77114
ID AAU77114 standard; Protein; 1255 AA.
AC AAU77114;
DT 05-JUN-2002 (first entry)
DE Human Her-2/neu polypeptide.
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
OS Homo sapiens.
XX W0200213847-A2.
XX 21-FEB-2002.
XX 13-AUG-2001; 2001WO-US25408.
XX 14-AUG-2000; 2000US-0638280.
XX 28-SEP-2000; 2000US-0675904.
XX (CORI-) CORIXA CORP.
XX Gaiger A, Cheever MA, Hand-zimmermann S;
XX WPI; 2002-280741/32.
XX N-PSDB; ABK10730.
XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
PS Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6728; DB 23; Length 1255;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPFGAASQTCTGDMKRLRPASPETHLDMRLHLYQCGVQVQGNL 60
DB 1 MELAALCRWGLLLALLPFGAASQTCTGDMKRLRPASPETHLDMRLHLYQCGVQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVROVPLRLRIVRGTOLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVROVPLRLRIVRGTOLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVTCASPGGLREQLRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
DB 121 DPLNNTTPTVTCASPGGLREQLRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLTRTCAGGCARCKGLPTDCCHEQC 240
DB 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLTRTCAGGCARCKGLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHNEQVTAEDGTQRCCKSKPCARVCYGLGMVYIKANSKFIGIT 360
DB 301 YNYLSTDVGSCTLVCPHNEQVTAEDGTQRCCKSKPCARVCYGLGMVYIKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPELOVFTLEBITGYLISAWPDSLP 420
DB 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPELOVFTLEBITGYLISAWPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYS LTLQGLIGISWGLRSRLGSLALIHNNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYS LTLQGLIGISWGLRSRLGSLALIHNNTHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSOFURGQSC 540
DB 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSOFURGQSC 540
QY 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600
DB 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600
QY 601 PSGVKPDLSYMPIWKPFDEBAGACQPCINCTHSCVDLDDKGCFAEQRASPLTISVAVVG 660
DB 601 PSGVKPDLSYMPIWKPFDEBAGACQPCINCTHSCVDLDDKGCFAEQRASPLTISVAVVG 660
QY 661 ILLVVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQWRIILKETEL 720
DB 661 ILLVVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQWRIILKETEL 720
QY 721 RKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVSP 780
DB 721 RKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVSP 780
QY 781 YVSRLLGICLTSTVOLVTLMPYGCCLLDHVRENRLGSGDQLLNWCMQIAKMSYLEVDV 840
DB 781 YVSRLLGICLTSTVOLVTLMPYGCCLLDHVRENRLGSGDQLLNWCMQIAKMSYLEVDV 840
QY 841 LVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRFT 900

Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRRPT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPDPPICTIDVTMIMVKWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPDPPICTIDVTMIMVKWM 960
QY 961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020
Db 961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020
QY 1021 BEYLVPQOGFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080
Db 1021 BEYLVPQOGFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080
QY 1081 AGSDVFDGLGMAAGLQSLPTHDPSPLOQYSEDPTVPLSETDGYVAPLTCSPQPEV 1140
Db 1081 AGSDVFDGLGMAAGLQSLPTHDPSPLOQYSEDPTVPLSETDGYVAPLTCSPQPEV 1140
QY 1141 NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVPAFGGAVENPEYLTPO 1200
Db 1141 NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVPAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 15
AA39568
ID AAR39568 standard; Protein; 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN W09316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU) CETUS ONCOLOGY CORP.
XX
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston LL, Huston JS, Oppermann H, Ring DB;
XX
XX WPI; 1993-272889/34.
DR N-PSDB; AAQ46083.
XX

New single chain Fv polypeptide binding to C-erbB-2 tumour antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;
Query Match 98.1%; Score 6685; DB 14; Length 1433;
Best Local Similarity 98.1%; Pred. No. 0;

Matches 1231; Conservative 7; Mismatches 17; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPETHLDMRLHYQCCVQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPETHLDMRLHYQCCVQVQGNL 60
QY 61 ELTYLPTNASISFLQDIOEVQGYVLIHNRQVRQVPLQRLIRVRGTQOLFEDNYALAVLNG 120
Db 61 ELTYLPTNASISFLQDIOEVQGYVLIHNRQVRQVPLQRLIRVRGTQOLFEDNYALAVLNG 120
QY 121 DPLNNTTPTVGTASPGGLRELOLRSLTEILKGGVLIQRPOLCYQDTILWKDIFHKQNQLA 180
Db 121 DPLNNTTPTVGTASPGGLRELOLRSLTEILKGGVLIQRPOLCYQDTILWKDIFHKQNQLA 180
QY 181 LTLDTNRSRACHPCSPMKSCWGESSEDCQSILTRTVACGGCARCKGPLPTDCCHEOC 240
Db 181 LTLDTNRSRACHPCSPMKSCWGESSEDCQSILTRTVACGGCARCKGPLPTDCCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
QY 301 YNYLSTDVGSCTLVCPHNVETADGTQRCCKSPCARVCYGLGMOYIKANSFPGIT 360
Db 301 YNYLSTDVGSCTLVCPHNVETADGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 360
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLBEITGLYVLSAMPDUL 420
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLBEITGLYVLSAMPDUL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRLSRLGSLGIALIHNHNLFCVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRLSRLGSLGIALIHNHNLFCVHTV 480
QY 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCGPGTQCVNCSQFLRGQSC 540
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCGPGTQCVNCSQFLRGQSC 540
QY 541 VEECRVLQGLPREVYNARHCLPCHPECOQNGSVTCFGEADOCVACAHYKDPPEVCVRC 600
Db 541 VEECRVLQGLPREVYNASHCLPCHPECOQNGSVTCFGEADOCVACAHYKDPPEVCVRC 600
QY 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTISIVSAVVG 660
Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTISIVSAVVG 660
QY 661 ILLVVVLGVVFGIILKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGIILKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVGS 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVGS 780
QY 781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRLGSLQDILLNMCQIAKMSYLEDVYR 840
Db 781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRLGSLQDILLNMCQIAKMSYLEDVYR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRRPT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRRPT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPDPPICTIDVTMIMVKWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPDPPICTIDVTMIMVKWM 960
QY 961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020
Db 961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020
QY 1021 BEYLVPQOGFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080
Db 1021 BEYLVPQOGFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080

Qy	1081	AGSDVFDGDLGMGAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMGAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVRFQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVFAFGGAVENPEYLTPO	1200
Db	1141	NQPDVRFQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVFAFGGAVENPEYLTPO	1200
Qy	1201	GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVAV	1255
Db	1201	GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVAV	1255

Search completed: July 22, 2003, 08:41:12
Job time : 43.9774 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-325-339-12
Perfect score: 6814
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6217	91.2	1259	6 O18735	O18735 canis famil
2	3144	46.1	1209	11 Q9QX70	Q9QX70 rattus norv
3	3115	45.7	1210	11 Q9EP98	Q9EP98 mus musculu
4	2727	40.0	1165	13 Q9YH40	Q9YH40 xiphophorus
5	2694	39.5	1137	13 Q9W6F6	Q9W6F6 gallus gall
6	2297	33.7	1328	13 P79754	P79754 fugu rubrip
7	2029	29.8	1433	5 Q9BIH9	Q9BIH9 anopheles g
8	1871	27.5	419	4 Q9UK79	Q9UK79 homo sapien
9	1739	25.5	367	11 Q8R2X1	Q8R2X1 mus musculu
10	1720	25.2	729	15 Q86712	Q86712 avian rous-
11	1718	25.2	567	15 Q86714	Q86714 avian rous-
12	1697	24.9	412	4 Q8WYV0	Q8WYV0 homo sapien
13	1653	24.3	962	15 Q64895	Q64895 avian eryth
14	1645	24.1	545	15 Q85468	Q85468 avian eryth
15	1506	22.1	655	11 Q9WVF5	Q9WVF5 mus musculu
16	1490	21.9	643	11 Q9ERV6	Q9ERV6 mus musculu

17 1276 18.7 1193 5 Q9YIX8
18 1190.5 17.5 1368 5 Q23821
19 1162 17.1 1717 5 Q26566
20 1126 16.5 527 13 Q90836
21 1001.5 14.7 478 11 Q9ES0
22 942.5 13.8 599 13 Q9PSH2
23 906 13.3 165 4 Q14256
24 887 13.0 176 11 Q923V5
25 806.5 11.8 346 13 P11776
26 778 11.4 435 5 Q8SZM1
27 754.5 11.1 311 13 Q9P162
28 734.5 10.8 1362 13 Q9PV24
29 734 10.8 331 4 Q9BUD7
30 730 10.7 1671 5 Q9NJV5
31 723 10.6 149 6 Q9BG66
32 692 10.2 1418 13 Q93457
33 686.5 10.1 1368 13 Q8UM85
34 671.5 9.9 1369 13 Q8UM86
35 662.5 9.7 1472 5 Q9U5A8
36 658 9.7 1412 13 Q8UM84
37 657 9.6 1358 13 Q73798
38 641.5 9.4 1418 13 Q8UM83
39 632 9.3 1245 13 Q9YGH8
40 622 9.1 1371 11 Q9QW4
41 620.5 9.1 2144 5 Q9VD94
42 598 8.8 987 11 Q91YMO
43 595 8.7 935 4 Q96L35
44 595 8.7 987 11 Q99MR2
45 587.5 8.6 1036 4 Q07912

ALIGNMENTS

RESULT 1
O18735
ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE ErBB-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbB-2 from canine mammary gland";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1; -
DR HSP; P11362; IFGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP_2_domain; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match		91.2%; Score 6217; DB 6; Length 1259;
Best Local Similarity		90.9%; Pred. No. 0;
Matches 1145; Conservative 43; Mismatches 66; Indels 6; Gaps 2;		
QY	1	MELAAACWGLLLALLPFGAASQTCTGDMKRLRASPETHDMLRHLQYQCQVVOGNL 60
DB	1	MELAAACWGLLLALLPFGAAGTCTGDMKRLRASPETHDMLRHLQYQCQVVOGNL 60
QY	61	ELTYLPTNASLFLQDIOEVQYVLI AHNOVQVPLQRLRI VRGTQLPEDNVALAVLNG 120
DB	61	ELTYLPANASLFLQDIOEVQYVLI AHNOVQVPLQRLRI VRGTQLPEDNVALAVLNG 120
QY	121	DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHRKNQLA 180
DB	121	DPLEGGIPAPGAAOGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHRKNQLA 180
QY	181	LTLIDNRSRACHPCSPCKSGRCSGESSEDCSLTRIVCAGGCARCKGPLPTDCCHQC 240
DB	181	LTLIDNRSRACHPCSPCKSGRCSGESSEDCSLTRIVCAGGCARCKGPLPTDCCHQC 240
QY	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTSCP 300
QY	301	YNYLSTDVSGSLVCLPHNQEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360
DB	301	YNYLSTDVSGSLVCLPHNQEVTAEDGTORCEKSKPCARVCYGLGMHLREVRAVTSAN 360
QY	361	ELEFAGCKKIFGSLAFLPESFGDDPASNTAPLOPEQLQVFETLEBITGYLISAWPDSLP 420
DB	361	IOEFAGCKKIFGSLAFLPESFGDDPASNTAPLOPEQLRVFEALBEITGYLISAWPDSLP 420
QY	421	DLVSFQNLQVIRGRLHNGAYSLTQGLGISWGLRSRLRELGLALIHNTLHCFVHTV 480
DB	421	NLSVFQNLQVIRGRLHNGAYSLTQGLGISWGLRSRLRELGLALIHNRNLFCVHTV 480
QY	481	PMQDLFRNPHQALLHTANRPECEGVEGLACHQLCARGHCWPGPTQCVCNCSQFLRGQEC 540
DB	481	PMQDLFRNPHQALLHSANRPEECVGEGLACYP-CAHGHWCWPGPTQCVCNCSQFLRGQEC 539
QY	541	VEECRVLOGLPREYVNAHCLCHPECPQNGSVTCFGEADQCACAHYKDPDFCVARC 600
DB	540	VEECRVLOGLPREYVNDKRYCLPCHSECQPNQSVTCFGEADQCACAHYKDPDFCVARC 599
QY	601	PSGVKPDLSYPMIWKFPDDEGACQPCINCHSCVDLDDKGPAPQASPLTSIVSAVVG 660
DB	600	PSGVKPDLSYPMIWKFADEEGTCQPCINCHSCADLDEKGPAPQASPLTSIIAAVVG 659
QY	661	ILLVVVLGVVFGILIKRRQOKIRKVTMRRLQETELVEPLTPSGAMPNQAQRILKETEL 720
DB	660	ILLAVVVLGVVFGILIKRRQOKIRKVTMRRLQETELVEPLTPSGAMPNQAQRILKETEL 719
QY	721	RKVVLGSGAFQTVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780
DB	720	RKVVLGSGAFQTVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 779
QY	781	VYSRLGLICTSTVOLVTLQMPYGLLDHVRNRCGLSQDLNNCMQIAKMSYLEVDV 840
DB	780	VYSRLGLICTSTVOLVTLQMPYGLLDHVRNRCGLSQDLNNCMQIAKMSYLEVDV 839
QY	841	LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
DB	840	LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIPPRFT 899
QY	901	HQSDVMSGVTVWELMTFGAKYDGI PAREIPDLLEKGERLPQPICTIDVYMWKCM 960
DB	900	HQSDVMSGVTVWELMTFGAKYDGI PAREIPDLLEKGERLPQPICTIDVYMWKCM 959
QY	961	IDSECRPRFRELVSFSESMARDPQRFVVIQNEDELGPASPLDSTFFVRSILLEDDMDGLVDA 1020
DB	960	IDSECRPRFRELVAEFSMARDPQRFVVIQNEDELGPASPLDSTFFVRSILLEDDMDGLVDA 1019
QY	1021	EYLVPPQGGFFCPDPAPGAGMGMVHRRSSSTRSGGDLTLGLEPSEEA PRSLAPSEG 1080
DB	1020	EYLVPPQGGFFCPBPTCAGGTAAHRRSSSTRSGGDLTLGLEPSEEA PRSLAPSEG 1079
QY	1081	AGSDVFDGDLGMAKGLQSLPHTDPSPLQRYSEDPVLPSPSETDGYVAPLTCSPQPEYV 1140
DB	1080	AGSDVFDGDLGMAKGLQSLPSPQSPPLQRYSEDPVLPSPSETDGYVAPLTCSPQPEYV 1139
QY	1141	NQPDVRPOPSPPREGPLPAARPAAGATLER-----AKTILSPGKNGVVKDVFATGGAVENPE 1195
DB	1140	NQPEVWPFPPLALEGPLPSPAGATLERPKTILSPGKNGVVKDVFATGGAVENPE 1199
QY	1196	YLTPOGGAAPHPHPPAFSPAFDNLVYWDQDPPERGAPPSTPKGTPTAENPEYLGLDV 1255
DB	1200	YLAPGRAAPHPHPPAFSPAFDNLVYWDQDPSERGSPPSTPEGTPTAENPEYLGLDV 1259
RESULT 2		
Q90X70		PRELIMINARY; PRT: 1209 AA.
ID	Q90X70;	
AC	Q90X70;	
DT	01-MAY-2000 (TRENBLrel. 13, Created)	
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)	
DT	01-JUN-2002 (TRENBLrel. 21, Last annotation update)	
DE	Epidermal growth factor receptor.	
GN	EGFR.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RX	MEDLINE=9025888; PubMed=2342466;	
RA	Petch L.A., Harris J., Raymond V.W., Blaeband A.J., Lee D.C.,	
RA	Earp H.S.;	
RT	"A truncated, secreted form of the epidermal growth factor receptor is	
RT	encoded by an alternatively spliced transcript in normal rat tissue.";	
RL	Mol. Cell. Biol. 10:2973-2982(1990).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Petch L.A.;	
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Guttridge K., Dawson T.L., Earp H.S.;	
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; M37394; AAF14008.1; .	
DR	HSSP; P11362; 1FGK	
DR	InterPro; IPR000494; EGFR_L domain.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001245; Tyr_pkinase.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep_L domain; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; TyRK; 1.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.	
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.	
KW	ATP-binding; Receptor; Transferrase; Tyrosine-protein kinase.	
SQ	SEQUENCE 1209 AA; 134891 MW; 96PFE7F6CC1B7773 CRC64;	
Query Match		46.1%; Score 3144; DB 11; Length 1209;
Best Local Similarity		50.3%; Pred. No. 3.1e-228;
Matches 643; Conservative 165; Mismatches 354; Indels 116; Gaps 28;		
QY	3	LAALCRWGLLLALLPFGA-ASTQVCTGDMKRLRASPETHDMLRHLQYQCQVVOGNL 61

Ds 15 LAALCAAG-----GALEKKVCGQSTNRLTQLGTFTEDHFLSLQRMNCEVVLGNLE 66
Qy 62 LTYLPTNASLFLQDIEQVGYVLIHNRQVPLQLRLIRVGTOLFDNVALAVLDNGD 121
Ds 67 ITTYQVNVYDLSTKTIQEVAGYVLIATNVERIPLENLOIRGNALYENTVALAVLSN-- 124
Qy 122 PLNNTPTVTGASPGRLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFHKNNQAL 181
Ds 125 -----YGTNKTGLRPLRNLOEILIGAVRFESNNPILCNMETIQRDIV-QDVFLSN 175
Qy 182 TLIDNRS-RACHPCSPCKSGRCWGESSEDCOSLTRTVCAAGCA-RCKGPLPTDCHEQ 239
Ds 176 MSMDVQRHLTCGCKDPCSPNGSCWGRGEENCOKLTKIICAOQCSRRRCGRSPSCCHNQ 235
Qy 240 CAAGCTGPKHSDCLACHFNHSGICELCPALVYNTDTTFESMPNPEGRYTFGASCVTAC 299
Ds 236 CAAGCTGPRSDCLVCHFRDEATCKDTCPLMLYNTTYQMDVNPCKYSGFATCVKCK 295
Qy 300 PYNVLTVDVGSCTLVCLPHNQVTAEDGTQCEKSKFCARCYGVGLGMOYTKANSKFTGI 359
Ds 296 PRNVVYTDHGSVCRAAGPDYEV-BEDGVSKCKCDGPRKVCNGIGIGEFK-DTSLINA 353
Qy 360 TELE-FAGCKKIFGSLAFLPESFGDPASTAPLOEQLOVFETLEETGLVLYISAWPDS 418
Ds 354 TNIKFKYCTAISGLHLPLVAFKGDSTRTPLDPRELEILKTKVKEITGELLQAWPEN 413
Qy 419 LPDLSVFQNLQVIRGRILHNGAYSILTQGLGSLMLGSLRSLRELGSGLAIHNNHLCFVH 478
Ds 414 WTDLHAFENLIIIRGRTHQGFSLAVVGLNITSLGSLSLKXISGDVLIISGRNLCVAN 473
Qy 479 TVPMDQLFRNHQALLHTANREBECVGEGLACHOLCARGHGWGPTQCVNCSOFLRQ 538
Ds 474 TINKKLFGTGNQTKIMNNAEKDCKATNHCNPLCSSEGCWGPDEPTDCVSCQNVSRGR 533
Qy 539 ECVEECRVQLGPREYVNAHCLPCHPECQPNQSVTCFGEADOCVACAHYKDPFCVA 598
Ds 534 EVDKCNLEGEPRFVENSCTIOCHPECLPOTMNTCTGRGPDNCKCAHYVDGPHCVK 593
Qy 599 RCPGSKVPLDLYMPTKPEDEGACQPCPINCTHSCVDLDKGCPEAQASPLTSIVSA 657
Ds 594 TCPSGINGENNTL-VKMFADANNVCHLCHANCYGCAGPLKGC--QQPEGPKPSIATG 650
Qy 658 VGIILVVVLGVFGI-LIKRQOKIRKYWRRLLOETELVEPLTPSGAMPNQAMRLK 716
Ds 651 IVGGLLFTVW-VALGIGLFMRRLQVRLRRLRLQLRELVEPLTPSGEAPNQHLRLK 709
Qy 717 ETELRKVKVLGSGAGFTYVGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAG 776
Ds 710 ETEFKIKVLGSGAGFTYVGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAS 769
Qy 777 VGSYVSRLLGICLTSTVQLTQMPYGLLDHVRNENRGRGLGSDLLNMCQIAKMSYL 836
Ds 770 VDNPHVCRLLGICLTSTVQLTQMPYGLLDHVRNENRGRGLGSDLLNMCQIAKMSYL 829
Qy 837 EDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIMWALESLR 896
Ds 830 EDRELVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIMWALESLR 889
Qy 897 RRFTHQSDVSGYVTVWELMTFGAPYDGIAPAREIPDLLEKGERLPQPICTIDVYMTW 956
Ds 890 RIYTHQSDVSGYVTVWELMTFGAPYDGIAPAREIPDLLEKGERLPQPICTIDVYMTW 949
Qy 957 KCMWIDSECRPRFRELSEFSEMRARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMG 1015
Ds 950 KCMWIDSECRPRFRELSEFSEMRARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMG 1009
Qy 1016 DLVDAEYLVPOQGFCCPDPAAGAGVMVHRHRSSTSSRSGGDLTLGLPESEEEAPRSL 1075
Ds 1010 DVWDAEYLVPOQGF-----NSPST-----SRTPL 1035
Qy 1076 APSEGACSDVFDGLGMAKGLQSLPHDPSPLPQVSEDTVPLPST--DGVVAPLTC 1133
Ds 1036 LSSLSANSN-----SSTVACINRNGSCRVKEDAFILQRYSSDPTSVLTEDNIDDTFL----- 1086

Qy 1134 SPQEVYVNPQDPVRPQPPSPREQPLPAAPAGATLRAKTLSPGKNGVVKDVFAGGAVEN 1193
Ds 1087 -PVPEYINQ-SVPKEPAGSVQNPVYHNPQPLHP-----APGRDLHYQN--PHSNAVSN 1134
Qy 1194 PEYL-TPQGAAPQHPPPAFSPAFDNLYYDQ-----DP-----PERGAPPST 1236
Ds 1135 PEYLNTAQ-----PTCLSSGFDSSALMTQKSHQMSLDNPDYQDFFPKPEAKPNGI 1185
Qy 1237 FKGTPTAENPEVGLDVP 1254
Ds 1186 FKGTPTAENPEVGLDVP 1202
RESULT 3
QSEP98
ID QSEP98 PRELIMINARY; PRT: 1210 AA.
AC QSEP98;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/svJ, AND 129/SVEVTC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAG28045.1;
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG28045.1;
DR HSSP; P11362; 1FGK.
DR MGD; MGI:95294; Egr.
DR InterPro; IPR000345; CytC heme bind.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

KW	ATP-binding; Receptor; Transferase.	
SQ	SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;	
	Query Match 45.7%; Score 3115; DB 11; Length 1210;	
	Best Local Similarity 49.8%; Pred. No. 4.8e-226;	
	Matches 634; Conservative 167; Mismatches 360; Indels 112; Gaps 25;	
QY	11 LLLALLPPGAA--STQVCTGTDMKRLRASPETHLMLRHLVQGVQVQGNLELYPTN 68	
DB	14 LLTALCAAGGALEKKVCGQTSNRLTQLGTFEDHFLSLQRMVYNNCEVILGNLEITYVQRN 73	
QY	69 ASLSFLQIDQVGVVLIHNOVRVPLQRIVRGTLQFEDNYALAVLDNGDPLNNTTP 128	
DB	74 YDLSFLKTIQEVAGVYLIANTVERIPLENLQIRGNALYENTYALAILSN----- 124	
QY	129 VTGASPGGLRELQLBSLLEILKGGVLIQRNPOLCYQDTILWKDI----PHKNQLALILI 184	
DB	125 -YGTNRTGLRELPMNLQELIGAVRFSPNNPILCNNDTIQMRDIVONVFMNSMDL--- 180	
QY	185 DTRSRACHPCSPMCKGSRWCSESSEDQSLTRTRVCAGGCA-RCKGPLPTDCCHEQCAAG 243	
DB	181 -QSHPSCKPCDPCPNSCGWCGGEGENCQKLTIIQAQCSHRCGRSPSDCCHNQCAAG 239	
QY	244 CTGPKHSDCLACLNHNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACPNY 303	
DB	240 CTGPRESCLVCQKQDEATCKDTCPLMLYNPTTYQMDVNPGEKYSFGATCVKCKPRNY 299	
QY	304 LSTDVGSCTLVCPLHNQVETADGTQCEKSKPCARVCYGLGMQYIKANSKFIGITELE 363	
DB	300 VYTHGSCURACGPPYEV-EDGIRKCKKCGKRCVNGIGIGEFK-DTLISINATNIK 357	
QY	364 -FAGCKKIFGSLAPLPEFDGDPASNTAPLOEQLOVQFETLEETCYLYISAWPDSLPDL 422	
DB	358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWDL 417	
QY	423 SVFQNLQVIRGRILNNGAYSLTLOGIGISWGLRSLRELGSGLALIHNTLHCFVHTVPM 482	
DB	418 HAFENLEITIRGTHQKQFSLAVVGLNITSLGSLKEISDGDVITISGRNRLCYANTINW 477	
QY	483 DQLFNPHQALLHTANRPEDECVCGEGLACHOLCARHGWCGPGTQCVNCSQFLRGQECVE 542	
DB	478 KKLFTGPNQTKIMNNAEKCKKAVNVCNPLCSGEGCWGPEPROCVSQNVSRGECVE 537	
QY	543 ECRVLQGLPREYVNAHCLPCHPEQOPQNGSVTCFGEADQCACAKHDPDFCVARCP 602	
DB	538 KCNILEGEPRFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597	
QY	603 GVKPDLSTWPIWKFPDEGACQPCPNCTHSCVDLDDKGCAPAEQASPLTSIVSAVGIL 662	
DB	598 GIMGENNTL-VNKYADANNVCHLCHANCTYGCAGFLQGCVEVWPSGPKIPSIATGIVGGL 656	
QY	663 LVVVLGVVFGI-LIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAOBRIKLETBLR 721	
DB	657 LFIVV-VALGIGLFWRRHIVKRTLRLLQRELVEPLTPSGEAPNAHLKILKETBFK 715	
QY	722 KVKVLGSAFGTVYKGIWIPGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGS 781	
DB	716 KIKVLGSAFGTVYKGLWIPGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPH 775	
QY	782 VSRLLIGLCTSTVQLTQVLMFYGCLLDVHRENRGRGLSGQDLLNWCNQIAGKMSYLEDVRL 841	
DB	776 VCRLLIGLCTSTVQLTQVLMFYGCLLDVYREHKDNIGSYQLLNWCVQIAGKNYLEDRL 835	
QY	842 VHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNMALESILRRRTTH 901	
DB	836 VHRDLAARNLVKTPQHVKITDFGLAKLLGAEKEYHAEKGVKPKKNMALESILHRIYTH 895	
QY	902 QSDVMSYGVTVWELMTFAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVVMIMVKCMI 961	
DB	896 QSDVMSYGVTVWELMTFGSKPYDGIIPASDISILEKGERLPQPPCTIDVVMIMVKCMI 955	
QY	962 DSECRPRELVSFESRNARDPQRVWLTQ-NEDLGPASPLSTFVRSLLDDDDGDLVDA 1020	

DB	956 DADSRPKFRELILEPSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMEDVDA 1015	
QY	1021 EYLIVPOQGFPCPDAPGAGGVMHHRSSSTRSGGDLTLGLSESEEAEPRLAPSEG 1080	
DB	1016 DEYLTPOQGGFF-----NSPST-----SRTPLLSLS 1041	
QY	1081 AGSDVFDGDLGMAAKGLQSLPTHDPSPQLQRYSEDTVPLPSET--DGYVAPLTCSPQPE 1138	
DB	1042 ATSN-----NSTVACINRNGSCRVKEDAFQRYSSDPTGAVTEDNIDDAFL-----PVPE 1091	
QY	1139 YVNPQDVPRQPPSPREGPLPAARPAAGATLERAKTLPSPGKGVVQDVAFGGAVENPEYL- 1197	
DB	1092 YVNO-SVPKRPAGSVQNPVYHNOPLRP-----APGRDLHYQN--PHSNAVGNPEYLN 1140	
QY	1198 TPQGAARQPPPPPAFPAFONLYYNDQ-----DP-----PERGAPSTPKGTP 1241	
DB	1141 TAO-----PTCLSSGFNSPALMIQKSHQMSLDNPDYQODFFPKETKNGIFKG-P 1190	
QY	1242 TAENPEYLGLDVP 1254	
DB	1191 TAENAEYLRVAPP 1203	
RESULT 4		
QYH40	ID	Q9YH40 PRELIMINARY; PRT; 1165 AA.
AC	Q9YH40;	
DT	01-MAY-1999 (TRENBLrel. 10, Created)	
DT	01-OCT-2000 (TRENBLrel. 15, Last sequence update)	
DT	01-JUN-2002 (TRENBLrel. 21, Last annotation update)	
DE	Receptor tyrosine kinase proto-oncogene.	
GN	XMRK.	
OS	Xiphophorus xiphidium.	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;	
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;	
OC	Cyprinodontiformes; Poeciliidae; Xiphophorus.	
OX	NCBI_TaxID=8086;	
RN	(1) SEQUENCE FROM N.A.	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=RIO PURIFICATION;	
RX	MEDLINE=98241172; PubMed=9582016;	
RA	Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,	
RA	Altmann J., Schartl M.;	
RT	"Activation of the Xmrk proto-oncogene of Xiphophorus by	
RT	overexpression and mutational alterations.";	
RL	Oncogene 16:1681-1690(1998).	
RN	(2) SEQUENCE FROM N.A.	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=RIO PURIFICATION;	
RA	Schartl M.;	
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; U53471; AADI0500.2; -;	
DR	HSSP; P11362; 1FGK.	
DR	InterPro; IPR000345; CytC heme bind.	
DR	InterPro; IPR000494; EGFR_L domain.	
DR	InterPro; IPR000719; Euk_Pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001899; Gram_pos_anchor.	
DR	InterPro; IPR001245; Tyr_pkinase.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF01030; Recep_L domain; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; TyrKc; 1.	
DR	PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.	
DR	PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.	
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 2.	
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.	
KW	ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.	

[illegible]

RESULT 7

QY	378	PESFDGDPASNT-----APLQPEOLQVFETLEETLGTLYTISAWPDSLPLDLSVFONLQ	429
Db	330	DQSFDFGQQVYTFNPSFGPRVYKIDPDLRLEVFSTVFKEITGFINAQHHNPTFTTLYNFRLNE	389
QY	430	VIRGRILHNGAY-SLTLQGLGISWGLGRLSRLREISGSLALIHNNTHLCFVHTVPMWQDLFRN	488
Db	390	VVGRQLKENLFPASVIYKTSLSKLSLELSLKRWNVSGSIVILENSDLCFVEDIDMSEIKKS	449
QY	489	PHQALLHTANRPEDEVCVGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQECVVEECRVLQ	548
Db	450	SDHEVMVQKNRNATECHEEGMEWESQCSKAGCKWGKPEQCLECNKVKYKGLUDSKCK---	506


```

Query Match      25.5%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 6.2e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

Qy 889 MALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICT 948
Db 1 MALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICT 60

Qy 949 IDVYIMVWKWMDISECRPRELSEFSESRWARDPQRVFIQNEGLGPASPLDSTFYRSL 1008
Db 61 IDVYIMVWKWMDISECRPRELSEFSESRWARDPQRVFIQNEGLGPASPLDSTFYRSL 120

Qy 1009 LEDDMGDLVDAEYLVPOQGFPCDDPAPGAGMWHHRSSSTRSGGGDLTLGLERSEE 1068
Db 121 LEDDMGDLVDAEYLVPOQGFPCDDPAPGAGMWHHRSSSTRSGGGDLTLGLERSEE 180

Qy 1069 EAPRSLAPSGAGSDVFDGLGCAAGLQSLPHDPSLQRYSEDTVPILPSETDGYV 1128
Db 181 EAPRSLAPSGAGSDVFDGLGCAAGLQSLPHDPSLQRYSEDTVPILPSETDGYV 240

Qy 1129 APLTCSPOEYVNPQVVRPPSPREGPLPAARPAAGATLERAKTILSPCKNGVVKDVFAFG 1188
Db 241 AFLACSPQEVYVNPQVVRPPSPREGPLPAARPAAGATLERAKTILSPCKNGVVKDVFAFG 300

Qy 1189 GAVENPEYLTQCGAPOPHPAPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEY 1248
Db 301 GAVENPEYLTQCGAPOPHPAPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEY 360

Qy 1249 LGLDVPV 1255
Db 361 LGLDVPV 367

RESULT 10
Q86712 ID Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnsson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSP; P03322; IAEs.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EPE1D63 CRC64;

Query Match      25.2%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 4.7e-121;
Matches 358; Conservative 118; Mismatches 118; Indels 102; Gaps 15;

```

```

Qy 569 PQNGSVTCFGPEADQCVACAHYKDPFCVARCPSGVKPDLSYMPYWKFPDEEGACQPCPI 628
Db 141 PEETATPKTP--DHCMKCAHFIDGPHCVKACPAAGVNGDNL-VWKYADANAVCOLCHP 197

Qy 629 NCTHSCVDLDDKGPAPORASPLTSIVSAV--GILLVVLVGVVFGILIKRROOKIRKYYTM 687
Db 198 NCTRGCKPGLEGCP--NGSKTPSIAAGVVGGLLCLVVGIGLIGLYLRR--HIVRKRTL 253

Qy 688 RLLQETELVPLTPSGAMPNQAOMRIILKETELKVKVVLGSGAGFCVTVYKGIWIPGENVK 747
Db 254 RLLQERELVPLTPSGEAPNQAHLRIILKETEFKVKVVLGSGAGFCVTVYKGLWIPGEKVK 313

Qy 748 IPVAIKVLRENTSIPKAKEIILDEAYVMAGVSPYVSRLLGLICTLTSTVOLVTLMPYGCIL 807
Db 314 IPVAIKELREATSPKAKEIILDEAYVMASVDNPRVCRLLGLICTLTSTVOLITQLMPYGCIL 373

Qy 808 DHVTENRGRGSLQDLLNWKMOIAKMSYLEDLVRLVHRLAARNVLVKSNNHVKITDFGLA 867
Db 374 DYIREHKDNIGSQYLLNMCVQIAKGMNYLEERLLVHRLAARNVLVKTPOHVKITDFGLA 433

Qy 868 RLLDIDETEHADGKVPKIMMALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGI 927
Db 434 KLLGADEKEYHAEKVPKIMMALESILHRIYTHOSDVMSYGVTVWELMTFGSKPYDGI 493

Qy 928 AREIPDLLEKGERLPQPPICTIDVYIMVWKWMDISECRPRELSEFSESRWARDPQREV 987
Db 494 ASEISSVLEKGERLPQPPICTIDVYIMVWKWMDISDRPKERELIAEFSKWARDPPRYL 553

Qy 988 VIQ-NEDLGPASPLDSTFYRSLLEDGMDGLVDAEYLVPOQGFPCDDPAPGAGMWHHR 1046
Db 554 VIQDERMHLPSPTDSKPYRTLMBEEDMEDIVDAEYLVPHQGF----- 598

Qy 1047 HRSSTRSGGDLTLGLEPSEEA PRSP-----APSEGAGSDVFDGLGMAAGLQSL 1101
Db 599 -NSPST-----SRTPLLSSLSATSNNSATNCID-----RNGOQH 631

Qy 1102 PTHDPSPLQRYSEDPTVPLPSET--DGVVAPLTCSPOEYVNPQVVRPPSPREGPLPA 1159
Db 632 PVREDSFQRYSSDPTGNFLSEIDDGFL-----PAPEYVNO--LMPKKS----- 675

Qy 1160 ARPAGATLERAKTILSPCKNGVVKDV-----AFGAVENPEYL 1197
Db 676 -----TAMVQNQIYNNISLTATSKLPMDSRYQNSHSTAVDNPEYL 715

RESULT 11
Q86714 ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnsson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.

```

```
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR Tyrosine-protein kinase.
FT NON TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match      25.2%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 4.6e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

Qy 578 GPEADQCACAHYKDPFCVACRSGVKPDLISYPIWKFPPDEGACQCPINCTHSCVDL 637
Db 1 GP--DHCWKCAHFIDGPHCVKACPAVLGENDTL-VKMYADANAVCOLCHPNCIRGCKGP 57

Qy 638 DDKCPABQASPLTSIVAVV-GILLVVVLGVGVGILIKRQOKIRKYTRRRLLQETEL 696
Db 58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVGGLGIGLYLRRR-HIVKRTLRLLQEREL 113

Qy 697 VEPLTPSGAMPNQAMRIKTELKVKVGLSGAGFTVYGIWIPDGENVKIPVAIKVL 756
Db 114 VEPLTPSGEAPNQHILKTEFKVKVGLSGAGFTVYGLMIPGEKVKIPVAIKEL 173

Qy 757 ENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVOLVTQLMPYCGLLDHYVRENKR 816
Db 174 EATSPKANKEILDEAYVMASVDNPRVCLLGICLTSTVOLITQLMPYCGLLDYIREHKN 233

Qy 817 LGSODLLNWCQIAKGSYLEDVRLVHRDLAARNVLYKSPHHVKITDFGLARLLDIDTE 876
Db 234 IGSQYLLNWCQIAKGMNLYLEERLVHRDLAARNVLYKTPQHVKITDFGLAKLLGADEKE 293

Qy 877 YHADGKVPKIMWALESILRRRFTHOSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLE 936
Db 294 YHAEGKVPKIMWALESILRIYTHOSDVWSYGVTVWELMTFGSKPYDGI PASEISSVLE 353

Qy 937 KGERLPQPICTIDVYIMVWCWIMIDSECRPRFRELVSFSEFMRDAPORFVVIQ-NEDLG 995
Db 354 KGERLPQPICTIDVYIMVWCWIMIDADSRPKFRELIAEFKSMARDPPRYLVIQDERMH 413

Qy 996 PASPLDSTFYSLLEDDMDGLVDAEYLVPOQGFCCPDPAAGAGWVHHRSSSTRSG 1055
Db 414 LPSPTDSKFTYRLTMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454

Qy 1056 GGDLTGLGLEPSEEEAPRSP-----APSEGAGSDVFDGLGMAAKGLQSLTPHDPSPLO 1110
Db 455 -----SRTPLLSSLSATSNATNCID-----RNCQGHVPVEDSFVQ 491

Qy 1111 RYSEDPTVPLPSET--DGYVAPLTCSPOPEYVNPQDVVRPQPPSPREGPLPAARPAATLE 1168
Db 492 RYSDPTGNFLEESIDGFL-----PAPEYVNQ--LMPKPKFS----- 526

Qy 1169 RAKTLSPKNGVVKDVP-----AFGGAVENPEYL 1197
Db 527 ----TAMVQNOIYNNISLTAKLPMDSRYQNHSHTAVDNPEYL 566

RESULT 12
Q8MYVO
ID Q8MYVO PRELIMINARY; PRT; 412 AA.
AC Q8MYVO;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN P31659.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
```

```
Wa D.P., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell growth."
RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF318349; AAL55856.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF HAND; UNKNOWN 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match      24.9%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 1e-119;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

Qy 889 MALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICT 948
Db 1 MALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICT 60

Qy 949 IDVTYIMVWCWIMIDSECRPRFRELVSFSEFMRDAPORFVVIQNEDELGPASPLDSTFYRSL 1008
Db 61 IDVTYIMVWCWIMIDSECRPRFRELVSFSEFMRDAPORFVVIQNEDELGPASPLDSTFYRSL 120

Qy 1009 LEDDDMDGLVDAEYLVPOQGFCCPDPAAGAGWVHHRSSSTRSGGDLTLGLEPSEE 1068
Db 121 LEDDDMDGLVDAEYLVPOQGFCCPDPAAGAGWVHHRSSSTRSGGDLTLGLEPSEE 180

Qy 1069 EAPRSLAPSGAGSDVFDGLGMAAKGLQSLTPHDPSPQRYSEDTVPLPSETDGVY 1128
Db 181 EAPRSLAPSGAGSDVFDGLGMAAKGLQSLTPHDPSPQRYSEDTVPLPSETDGVY 240

Qy 1129 APLTCSPOPEYVNPQDVVRPQPPSPREGPLPAARPAATLERAKTLSPKNGVVKDVFAG 1188
Db 241 APLTCSPOPEYVNPQDVVRPQPPSPREGPLPAARPAATLERAKTLSPKNGVVKDVFAG 300

Qy 1189 GAVENPEYLTPOGAAAPQ-----HPPPA---FSPAFLNL 1220
Db 301 GAVENPEYLTPOGAAALSP-----GAPPSTFKGTPTAEN 1245

Qy 1221 YYWD-QDPPER-----GAPPSTFKGTPTAEN 1245
Db 361 VWMTQCEPEQVRRSPDVSSGREGLTSAQIKRWEGPPTTSRGTCCHARN 410

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
DE Gag, v-erb-A, v-erb-B.
GN GAG, v-erb-A, v-erb-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
ON NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RA Bruskina A., Jackson J., Bishop J.M., McCauley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the transforming potential of the oncogene v-erb-B."
RL Oncogene 5:15-24 (1990).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -.
```

```
DR EMBL: X52211; CAA36459.1; JOINED.
DR HSSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk pkinese.
DR InterPro; IPR000536; Hormone rec.lig.
DR InterPro; IPR001723; Stchrnn_receptor.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinese; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDPINGER.
DR ProDom; PR000001; Euk_pkinese; 1.
DR ProDom; PR000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; TyrKc; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match      24.3%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 7.4e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 541 VEECRVLOGLPRE-YVNAR-HCLP-----CHPEQC 568
DB 354 IEKQESYLLAFEHYINRYKHNPHFWSKLLMKVADLRMIGAYHASRELFHMKEVCEPTELS 413
QY 569 PONGSVTCFGEADQCAVACHYKDPFCVAVCPGKPDLSVMPWKFPDEEGACQPCPI 628
DB 414 PQE-----VGP--DHCWKCAHFDGPHCVKACPAVLGENDTL-VWKYADANAVQCQLCHP 465
QY 629 NCTHSCVDLDDGCPAEQASPLTSIVSAV-GILLVVVLGVFGILIKRRQKIRKYTM 687
DB 466 NCTRGCKPGLEGCP---NGSKTPSIAAGVVGGLLVVGVGLGILYLR--HIVRKRTL 521
QY 688 RLLOETELVELPLTPSGAMPNQAMRIKTELKVKVKGSGAGFTVYKGIWIDGENVK 747
DB 522 RLLOERELVELPLTPSGEAPNQAHRLIKETEFKVKVKGSGAGFTVYKGLWIPEGEKVT 581
QY 748 IPVAIKVIRENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVOLVQLMPYGCLL 807
DB 582 IPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLITQLMPYGCLL 641
QY 808 DHVRENRLGSLQDLNMCQIAKGMYSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLA 867
DB 642 DYIREHKDNGSQYLLNMCVQIAKGMNLYEERHVMHRDLAARNVLKTPQHVKITDFGLA 701
QY 868 RLLOIDETEVHADGKVPKIMMALESILRRFTQSDVWSYGVTVWELMTGAKPYGIP 927
DB 702 KQLGADEKEYHAEGKVPKIMMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIP 761
QY 928 AREIPDLLEKGERLPQPPICITIDVYIMVCKWMDSECRPRFRELVSFSESMARDPQFV 987
DB 762 ASEISSVLEKGERLPQPPICITIDVYIMVCKWMSGASRPKRELIAEFSKWARDPPRYL 821
QY 988 VIO-NEDLGASPLDSTFYRSLLEDDMDGLVDABEYLVPQQGFPCDPAPGAGGMVHR 1046
DB 822 VIQDERMHLPSPTDSKYFRTLMEEDMEDIVDAEYLVPHQGF-----866
QY 1047 HRSSTSTRSGGDLTLGLEPSEEAAPRSLAPSEAGSDVDFCDLGMGAQKQLSLPTHDP 1106
DB 867 -NSPST-----SRTPLLSLSATS-----NSATKCIDRNGGH--898
QY 1107 SPLQRYSEDPTVLPSETDGVVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPGAT 1166
DB 899 -----PVREDGFL-----PAPEYVNO--LMPKPESTAMVQNIYNYISLT 936
```

```
QY 1167 -LERAKTSLSPGKNGVWVDVFAFGGAVENPEYL 1197
DB 937 AISKLPMDSRYQN-----SHSTAVDNPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
ID Q85468;
AC Q85468;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retrovirdae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CAA30024.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinese.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinese; 1.
DR ProDom; PD000001; Euk_pkinese; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DC88CCA0F8AF4 CRC64;
```

```
Query Match      24.1%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 1.4e-115;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 578 GPEADQCVACHYKDPFCVAVCPGKPDLSYMPWKFPDEEGACQPCINCTHSCVDL 637
DB 1 GP--DHCWKCAHFDGPHCVKACPAVLGENDTL-VWKYADANAVQCQLCHPNCTRGCKGP 57
QY 638 DDKCPAEQASPLTSIVSAV-GILLVVVLGVFGILIKRRQKIRKYTMRLLOETEL 696
DB 58 GLEGCP---NGSKTPSIAAGVVGGLLVVGVGLGILYLR--HIVRKRTLRLLOEREL 113
QY 697 VEPLTPSGAMPNQAMRIKTELKVKVKGSGAGFTVYKGIWIDGENVKIPVAIKVL 756
DB 114 VEPLTPSGEAPNQAHRLIKETEFKVKVKGSGAGFTVYKGLWIPEGEKVTIPVAIKE 173
QY 757 ENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVOLVQLMPYGCLLDHDVRENRL 816
DB 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLITQLMPYGCLLDYIREHK 233
QY 817 LGSODLLNMCQIAKGMYSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDE 876
DB 234 IGSODLLNMCVQIAKGMNLYEERHVMHRDLAARNVLKTPQDVKITDFGLAKQLGADE 293
QY 877 YHADGKVPKIMMALESILRRFTQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLE 936
DB 294 YHAEGKVPKIMMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSV 353
QY 937 KGERLPQPPICITIDVYIMVCKWMDSECRPRFRELVSFSESMARDPQFVVIQ-NEDLG 995
DB 354 KGERLPQPPICITIDVYIMVCKWMSDASRPKRELIAEFSKWARDPPRYLVIQDERMH 413
QY 996 PASPLDSTFYRSLLEDDMDGLVDABEYLVPQQGFPCDPAPGAGGMVHRSSSTRSG 1055
DB 414 LPSPTDSKYFRTLMEEDMEDIVDAEYLVPHQGF-----NSPST---454
```

```

QY 1056 GGLTLGLPSEEEAPRPL-----APSEGAGSDVFDGDLGMAKGLQSLPHTDPSPLQ 1110
Db 455 -----SRIPLSSLSATSNNSATNCIDRNG-----H----- 481
QY 1111 RYSEDPTVPLPSETDGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVNG--LMPKKPSTAMVQIQIYINVISLTAISK 523
QY 1170 AKTSLSPGNGVVKDVFARFGGAVENPEYL 1197
Db 524 LPMDSRYNQ-----SHSTAVDNPEYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor
DE isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maible N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ AND 129/SVEV7AC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner T., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -.
DR EMBL; AF275366; AAG28047.1; -.

```

```

DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -.
DR EMBL; AK004883; BAB23641.1; -.
DR EMBL; AK004911; BAB23662.1; -.
DR MGI; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.1%; Score 1506.5; DB 11; Length 655;
Best Local Similarity 44.5%; Pred. NO. 5.2e-105;
Matches 285; Conservative 97; Mismatches 233; Indels 25; Gaps 9;

QY 11 LLLALLPPGAA--STQVCTGTDMKLRUPASPTHLDMLRLHLYQGVQVQGNLELTYLPTN 68
Db 14 LTLTALCAGGAALEKKVCGTSTNLTQLGTTFEDHFLSLQRMYNCEVVLGNLEITYVQRN 73
QY 69 ASLSFLQDIQEVQGVYVLIHNAHQVRQVPLQRLIRIVRGTOFLFEDNLAVALDNGDPLNNTTP 128
Db 74 YLSFLTKTIQEVAGVLIALTNTVERIPLENQIIRGNALYENTYALALSN----- 124
QY 129 VTGASPGRLRELQRLSLTEILKGVLIQORNQOLCYQDTILMKDI----FHKNQLALTLI 184
Db 125 -YGTNRGTCLRELPMRLNQLQELIGAVRFNSNLTILCNMDTIQRWDIVQNVFMSNMDL--- 180
QY 185 DTRSRACHPCSPMKSGSRGSESEDCOSITRTVCAGGCA-RCKGPIPTCCCHCEQCAAG 243
Db 181 -QSHPSKCPKCDPSCPNGSCWGGGEENCQKLTIIAQCQSHRCRGRSPSCCHNQCAAG 239
QY 244 CTGPKHSCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPVNY 303
Db 240 CTGPRESCLVCKQFQDEATCKDTCPLMLNPTTYQMDVNPGEKYSFGATCVKCPRY 299
QY 304 LSTVGSCTLVCPILHNOEVTAEADGTQCEKSKPCARVCYGLGMOYIKANSKFIGITELE 363
Db 300 VVTDHGSQVACGPDYVEV-EEDGIRKCKKCDGPKRCVNGIGIGEFPK-DTLSINATNIK 357
QY 364 -FAGCKIFGSLAELPESFDGDPASNTAPLOEQLOVPETLEEITGYLYISAWPDSLPL 422
Db 358 HFKYCTAISGLDHLPLVAFKGSFTRTPPLDPRELEILTKVKEITGFLLIQAWPDNWTDL 417
QY 423 SVFQNLQVIRGRIHLNGAYSILTLQGLGISWLGSLRSLRELGLALIHHTHLCFVHTVPW 482
Db 418 HAFENLEIIRGTRKHQGFSLAVVGLNITSLGLSLKELISDGDVVISGNRLCYANTINW 477
QY 483 DQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVNCVSOFLRGOECVE 542
Db 478 KKLFGTNPQKTKIMNRAEKDKAVNHVCNPLCSSEGCGWGPEDRCDVSCQNVSRGECVE 537
QY 543 ECRVLQGLPREYVVARHCLPCHPCOPONGSVTCFGEADOCVACAHVKDPPFCVACRPS 602
Db 538 KCNLEGEPRFVENSECIQCHPECLPQAMMITCTGRGPDNCIOCAHYIDGPHCVKTCPA 597
QY 603 GVKPDLSPYMTWPKFDEBEGACQPCPINCTHSCVDLDDKGC 642
Db 598 GIMGENNTL-VWKYADANNVCHLCHNCTYCGAGPGLQGC 636

Search completed: July 22, 2003, 09:01:01
Job time : 54.3575 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-325-339-12
Perfect score: 6814
Sequence: 1 MELALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGIDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6728	98.7	1255	1	ERB2_HUMAN	P04626 homo sapien
2	5937	87.1	1257	1	ERB2_RAT	P06494 rattus norv
3	5921.5	86.9	1254	1	ERB2_MESAU	P00533 mesocricetu
4	3144	46.1	1210	1	EGFR_HUMAN	P00533 homo sapien
5	3118	45.8	1210	1	EGFR_MOUSE	Q01279
6	2966.5	43.5	1308	1	ERB4_HUMAN	Q15303 homo sapien
7	2947	43.2	1308	1	ERB4_RAT	O62956 rattus norv
8	2685.5	39.4	1167	1	XMRK_XIPMA	P13388 xiphophorus
9	2423.5	35.6	1342	1	ERB3_HUMAN	P21860 homo sapien
10	2351.5	34.5	1339	1	ERB3_RAT	O62799 rattus norv
11	1952	28.6	1426	1	EGFR_DROME	P04412 drosophila
12	1749.5	25.7	634	1	ERBB_ALV	P00534 avian leuko
13	1703	25.0	604	1	ERBB_AVIER	P00535 avian eryth
14	1630	23.9	540	1	ERBB_AVIEU	P11273 avian eryth
15	1595	23.4	703	1	EGFR_CHICK	P13387 gallus gall
16	1290	18.9	1323	1	LT23_CAEEL	P24348 caenorhabdi
17	1142.5	16.8	245	1	ERB2_MOUSE	P70424 mus musculu
18	727	10.7	1363	1	ILPR_BRALA	O02466 branchiocto
19	699	10.3	1300	1	IRR_MOUSE	Q9wt14 mus musculu
20	696	10.2	1383	1	INSR_RAT	P15127 rattus norv
21	695.5	10.2	1372	1	INSR_MOUSE	P15208 mus musculu
22	695	10.2	1382	1	INSR_HUMAN	P06213 homo sapien
23	693	10.2	1607	1	MIPR_LYMSY	Q25410 lymanaea sta
24	690	10.1	1297	1	IRR_HUMAN	P14617 homo sapien
25	685.5	10.1	1300	1	IRR_CAVPO	P14617 cavia porce
26	685	10.1	1477	1	HTK7_HYDAT	Q25197 hydra atten
27	641	9.4	1367	1	IG1R_HUMAN	P08069 homo sapien
28	623	9.1	1373	1	IG1R_MOUSE	Q60751 mus musculu
29	619.5	9.1	1370	1	IG1R_RAT	P24062 rattus norv
30	618	9.1	1390	1	INSR_AEDAE	Q93105 aedes aegypt
31	615	9.0	2146	1	INSR_DROME	P09208 drosophila
32	605	8.9	987	1	EPB4_HUMAN	P54760 homo sapien
33	599.5	8.8	984	1	EPB1_CHICK	Q07494 gallus gall

34	596.5	8.8	984	1	EPB1_RAT	P09759 rattus norv
35	590.5	8.7	977	1	EPB2_MOUSE	Q03145 mus musculu
36	590.5	8.7	984	1	EPB1_HUMAN	P54762 homo sapien
37	588	8.6	902	1	EPBB_XENLA	Q91736 xenopus lae
38	588	8.6	1114	1	RET_HUMAN	P07949 homo sapien
39	585	8.6	987	1	EPB4_MOUSE	P54761 mus musculu
40	584.5	8.6	976	1	EPB2_HUMAN	P29317 homo sapien
41	580.5	8.5	985	1	EPBA_XENLA	Q91571 xenopus lae
42	573.5	8.4	1053	1	PAK1_CHICK	Q00944 gallus gall
43	569	8.4	757	1	HT16_HYDAT	P53356 hydra atten
44	569	8.4	1068	1	FAK1_XENLA	Q91738 xenopus lae
45	563	8.3	1052	1	FAK1_MOUSE	P34152 mus musculu

ALIGNMENTS

RESULT 1
ERB2_HUMAN
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erbB-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993).
CC -/- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -/- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -/- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -/- SUBCELLULAR LOCATION: Type I membrane protein.


```
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 87.1%; Score 5937; DB 1; Length 1257;
Best Local Similarity 87.0%; Pred. No. 5.7e-308;
Matches 1094; Conservative 52; Mismatches 109; Indels 2; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASTVCTGTDMLRLPASPETHDMLRLHYQGCQVVGML 60
DB 1 MELAAWCRWGLLLALLPPGIAGTGVCTGTDMLRLPASPETHDMLRLHYQGCQVVGML 60
QY 61 ELTYLPTNASLSFLDIOIEVQGVVLIHQNQVQLRLIRIVRGTLQFEDNYALAVLDNG 120
DB 61 ELTYVPANASLSFLDIOIEVQGMVLIHQNQVQLRLIRIVRGTLQFEDNYALAVLDNR 120
QY 121 DPLNNTTPVT-GASPGGLRELQRLSLEILKGVLIQIRNPQLCYQDYLTKDIFKNNOL 179
DB 121 DQDNVAASTPGRTPEGLRELQRLSLEILKGVLIQIRNPQLCYQDYLTKDIFKNNOL 180
QY 180 ALTLDTNRSRACHPCSPMKSCRSWGESSEDCOSLIRTVGAGGACRCKGPLPTDCCHQ 239
DB 181 APVDITNRSRACPCAPACKDNHCWGESPEDCQLTGITCSGACRCKGRLPTDCCHQ 240
QY 240 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMNPNEGRTYFGASCVTAC 299
DB 241 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMNPNEGRTYFGASCVTAC 300
QY 300 PYNLYSTDVGSCTLVCPHLNQVBTAEDGTQRCCKSKPCARVCYGLGMQYIKANSKFIGI 359
DB 301 PYNLYSTEVSCTLVCPPPNQVETAEDGTQRCCKSKPCARVCYGLGMHRLGARAITSD 360
QY 360 TELEFAGCKKIFGSLAFIPESDGPASNTAPLOEQLOVFTLEITCYLYLISAWPDSL 419
DB 361 NVQEFDGCKKIFGSLAFIPESDGPSSGIAPLRPEQLQVFTLEITCYLYLISAWPDSL 420
QY 420 POLSVFQNLQVIRGRILHNGAYSLTLOGLGISWGLRLSRELGSGLALIHNNTHLCFVHT 479
DB 421 RDLVSFQNLRIIRGRILHDGAYSLTLOGLGIHSLGRLSRELGSGLALIHNAHLCFVHT 480
QY 480 VPWDQLFRNPQALLHTANRPEDE-CVSGEGLACHOLCARGHCWGPPTQVCNCSQFLRQ 538
DB 481 VPWDQLFRNPQALLHSGNREDIEDLCVSSGLVCNSLCAHGHWCWGPPTQVCNCSHFLRQ 540
QY 539 ECVEECRVLOGLPREVYNARHCLPCHPEQOPQNSVTCFPGPADOCVACHAKDPFCVA 598
DB 541 ECVEECRVWKGPREVYSDKRLCPCHPEQOPQNSSETCFGSADQCAACHYKDSSCVA 600
QY 599 RCPGSGVKPDLSPYMWKFPDEGACQPCPCINTHSCVDLDDKGCPAEORASPLTISVAV 658
DB 601 RCPGSGVKPDLSPYMWKFPDEGICQPCPCINTHSCVDLDERGCPAEQASPVTFIATV 660
QY 659 VGIILVVLGVVFGILIKRQOKIRKYTMRLQLQETELVEPLTPSGAMPNQAMRLKET 718
DB 661 VGVLLFLILVVVVGILIKRRQKIRKYTMRLQLQETELVEPLTPSGAMPNQAMRLKET 720
QY 719 ELRKVKVLGSGAFGVYGIWIPDGENVKIPVAIKVLRNTSPKANKEILDAYVMAGV 778
DB 721 ELRKVKVLGSGAFGVYGIWIPDGENVKIPVAIKVLRNTSPKANKEILDAYVMAGV 780
QY 779 SPVSRLLGICLTSTVQLVTQMPYGCLLDHRNRRGLSGDQLLNCWCMQIAGKMSYLE 838
```

```
DB 781 SPVSRLLGICLTSTVQLVTQMPYGCLLDHRNRRGLSGDQLLNCWCMQIAGKMSYLE 840
QY 839 VRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRR 898
DB 841 VRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRR 900
QY 899 FTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVC 958
DB 901 FTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVC 960
QY 959 WMIIDSECRPRFRELVSERFMRARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLV 1018
DB 961 WMIIDSECRPRFRELVSERFMRARDQRFVVIQNEIDLGPSSPMDSTFYRSLLEDGMDLV 1020
QY 1019 DAEYLYVPOQGFCDPDPAGGVMVHRHRSSTSGGDLTLGLEPSEEEAPRSLAPS 1078
DB 1021 DAEYLYVPOQGFSPDPTGPGTSTAHRHRSSTSGGDLTLGLEPSEEGPRSLAPS 1080
QY 1079 EGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQE 1138
DB 1081 EGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTLPLPETDGVVAPLTCSPQE 1140
QY 1139 VYNQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKVDFAFGGAVENPEYL 1198
DB 1141 YQGGAAPOPHPPAFSPAFDNLVYWDQPPPPGAPPSTFKGTPTAENPEYLGLDV 1255
QY 1199 FQGGAAPOPHPPAFSPAFDNLVYWDQPPPPGAPPSTFKGTPTAENPEYLGLDV 1257
DB 1201 PREGTASPPHSPASPAFDNLVYWDQNSSEGGPPSPNEGTPGTPTAENPEYLGLDV 1257

RESULT 3
ERB2_MESAU
ID ERB2_MESAU STANDARD; PRT; 1254 AA.
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

QY 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIILDEAYVMAGVGP 780
 Db 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIILDEAYVMAGVGP 780
 QY 781 YVSRLLGICLSTVOLVTLQMPYGCGLDHRVNRGRGLSGQDLLNMCQIAKMSYLEVDV 840
 Db 781 YVSRLLGICLSTVOLVTLQMPYGCGLDHRVNRGRGLSGQDLLNMCQIAKMSYLEVDV 840
 QY 841 LVHRDLAARNVLKSPNPKVITDFGLARLLDIDETEHADGKGKVPKMWALESIILRRFT 900
 Db 841 LVHRDLAARNVLKSPNPKVITDFGLARLLDIDETEHADGKGKVPKMWALESIILRRFT 900
 QY 901 HQSOWSVGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMIWVKCM 960
 Db 901 HQSOWSVGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMIWVKCM 960
 QY 961 IDSECRPRELVSFESRMARDPQRFVVIQEDLGPASPLDSTFYRSILLEDDMGDLVDA 1020
 Db 961 IDSECRPRELVSFESRMARDPQRFVVIQEDLGPASPLDSTFYRSILLEDDMGDLVDA 1020
 QY 1021 EYLVPQGGFFCPDPAPGAGGVHRRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080
 Db 1021 EYLVPQGGFFCPDPAPGAGGVHRRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080
 QY 1081 AGSDVDFDGLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Db 1081 AGSDVDFDGLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1141 NOPDVRPOPSPREGPLPAAPAGATLERAKTLSPGKGVKDVFAFGAVENPEYLTTPQ 1200
 Db 1141 NOPDVRPOPSPREGPLPAAPAGATLERAKTLSPGKGVKDVFAFGAVENPEYLTTPQ 1200
 QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 2
 ERB2_RAT
 ID -ERB2_RAT STANDARD; PRT; 1257 AA.
 AC P06494;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
 DE receptor-related protein).
 GN ERB2 OR NEU
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Neuroblastoma;
 RX MEDLINE=86118662; PubMed=3945311;
 RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
 RT "The neu oncogene encodes an epidermal growth factor receptor-related
 RT protein.";
 RL Nature 319:226-230 (1986).
 RN [2]
 RP SEQUENCE OF 852-905 FROM N.A.
 RC TISSUE=Sciatic nerve;
 RX MEDLINE=91222560; PubMed=2025425;
 RA Lai C., Lemke G.;
 RT "An extended family of protein-tyrosine kinase genes differentially
 RT expressed in the vertebrate nervous system.";
 RL Neuron 6:691-704 (1991).
 RN [3]
 RP STRUCTURE BY NMR OF 650-668.
 RX MEDLINE=92155181; PubMed=1346763;
 RA Gullick W.J., Bottomley A.C., Lofth F.J., Doak D.G., Mulvey D.,

RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
 RT "Three dimensional structure of the transmembrane region of the proto-
 RT oncogenic and oncogenic forms of the neu protein.";
 RL EMBO J. 11:43-48 (1992)
 CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
 CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X03362; CAA27059.1; ALT_INIT.
 CC PIR; A24562; TVRTNU.
 CC HSRP; P11362; 1FCG.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 3
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW proto-oncogene; Disease mutation.
 FT SIGNAL 1 21
 FT CHAIN 22 1257
 FT DOMAIN 22 654
 FT TRANSMEM 655 677
 FT DOMAIN 678 1257
 FT DOMAIN 159 369
 FT DOMAIN 473 646
 FT DOMAIN 722 989
 FT NP_BIND 728 736
 FT BINDING 755 755
 FT ACT_SITE 847 847
 FT DISULFID 196 205
 FT DISULFID 200 213
 FT DISULFID 221 228
 FT DISULFID 225 236
 FT DISULFID 237 245
 FT DISULFID 241 253
 FT DISULFID 256 265
 FT DISULFID 269 296
 FT DISULFID 300 312
 FT DISULFID 316 332
 FT DISULFID 335 339
 FT DISULFID 513 522
 FT DISULFID 517 530
 FT DISULFID 533 542
 FT DISULFID 546 562
 FT DISULFID 565 578

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC CC
DR EMBL: D16295; BAA03801.1; -.
DR HSSP: P11362; 1FGK.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR Pfam: PF02757; YLP_2_domain; 1.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; FU; 3.
DR SMART: SM00219; TyRK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21
FT CHAIN 22 1254
FT DOMAIN 22 652 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT TRANSMEM 653 675 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 676 1254 POTENTIAL.
FT DOMAIN 158 368 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 472 644 CVS-RICH.
FT DOMAIN 720 987 PROTEIN KINASE.
FT NP_BIND 726 734 ATP (BY SIMILARITY).
FT BINDING 753 753 ATP (BY SIMILARITY).
FT ACT_SITE 845 845 BY SIMILARITY.
FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 236 244 BY SIMILARITY.
FT DISULFID 240 252 BY SIMILARITY.
FT DISULFID 255 264 BY SIMILARITY.
FT DISULFID 268 295 BY SIMILARITY.
FT DISULFID 299 311 BY SIMILARITY.
FT DISULFID 315 331 BY SIMILARITY.
FT DISULFID 334 338 BY SIMILARITY.
FT DISULFID 511 520 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 531 540 BY SIMILARITY.
FT DISULFID 544 560 BY SIMILARITY.
FT DISULFID 563 576 BY SIMILARITY.
FT DISULFID 567 584 BY SIMILARITY.
FT DISULFID 587 596 BY SIMILARITY.
FT DISULFID 600 623 BY SIMILARITY.
FT DISULFID 626 634 BY SIMILARITY.
FT DISULFID 630 642 BY SIMILARITY.
FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 658 658 V -> E (IN ONCOGENIC NEU).
FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21P2B81 CRC64;

Query Match 86.9%; Score 5921.5; DB 1; Length 1254;
Best Local Similarity 86.6%; Pred. No. 3.8e-307;
Matches 1087; Conservative 61; Mismatches 106; Indels 1; Gaps 1;

Qy 1 MELAAACRWGLLLALLPFGAASSTVCTGTDMLRLPASPEHLDMLRHLGYCCQVQGNL 60
Db 1 MELAAACRWGLLLALLSFGASCTVCTGTDMLRLPASPEHLDIVRHLGYCCQVQGNL 60
```

QY 1141 NQPDVVRPQPPSPREGPLPAARAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTQ 1200
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 Db 1141 NQPEVRPQLTPEGFLPPVRPAGATLERKTLSPGKNGVVKDVFAGGAVENPEYLVPR 1200
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 QY 1201 GGAAPQPPPPAFSPAFDNLVYWDQDPGERGAPPSTFKGTPTAENPEYILGLDVPV 1255
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 Db 1201 GGSASQPH-PPALCPAFDNLVYWDQDPSESGSPPTFEGTPTAENPEYILGLDVPV 1254

RESULT 4
 EGFR_HUMAN
 ID EGFR_HUMAN STANDARD; PRT; 1210 AA.
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMDS; Q9UMG5; Q92795; O00732;
 AC O00688; Q9BZS2; Q9H2C9; Q9GXN1; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RL Nature 309:418-425(1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=95382957; PubMed=7654368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 epidermal growth factor receptor expression in human placenta.";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=97078686; PubMed=8918811;
 RA Reiter J.L., Maible N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 factor receptor gene encodes a truncated form of the receptor.";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=97256347; PubMed=9103388;
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 protein (TEGFR) in ovarian cancer.";
 RL Gynecol. Oncol. 65:36-41(1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RC TISSUE=Placenta;
 RX MEDLINE=21100872; PubMed=11161793;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Schell Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramanian S., Crossley T.D., Magnuson T.R., James C.D.,
 RA Maible N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 mouse alternative EGFR transcripts encoding truncated receptor
 isoforms.";
 RL Genomics 71:1-20(2001).
 RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schell C.M.,
 RA Lampland A.L., Balasubramanian S., Crossley T.O., Magnuson T.R.,

RA Maible N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 extracellular domain of the receptor.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Krueger W., Stolarek L.S., Weber W.,
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
 RT "Expression cloning of human EGF receptor complementary DNA: gene
 amplification and three related messenger RNA products in A431
 cells.";
 RL Science 224:843-848(1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 variety of RNAs overproduced in A431 carcinoma cells.";
 RL Nature 309:806-810(1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
 RA O'Malley B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 receptor cDNA from human A431 carcinoma cells.";
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
 RA Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 identification of sequences regulating its transcription.";
 RL Oncogene Res. 1:375-396(1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 transcript termination in the regulation of human epidermal growth
 factor receptor proto-oncogene RNA synthesis.";
 RL J. Biol. Chem. 266:1746-1753(1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 epidermal growth factor receptor gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN [13]
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RL Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor
 and supercoiled DNA.";
 RL Nature 309:270-273(1984).
 RN [15]
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
 RA Howk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor.";
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN [16]


```
QY 1141 NOPDVRPQPPSPREGPLPAARAGATLERAKTLSPGKNGVVKDVAFGGAIVENPEYL-TP 1199
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1094 NO-SYKPRAGSVQNPVYHQPLNP-----APSRDHYQD--PHSTAVGNPEYLVNTV 1142
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1200 QGAAPQPHPPAFSPAFONLYWYDQ-----DP-----PERGAPPSTFKGTPTA 1243
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1143 Q-----PTCVNSTFDSPAHWAQKSHQISLDNPYQODFFPKPEAKPNGIFKGS-TA 1192
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1244 ENPEYL 1249
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1193 ENAEYL 1198
||| : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
ID EGFR MOUSE STANDARD; PRT; 1210 AA.
AC Q01279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (BC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
(bek/KGFR) gene.";
RL Oncogene 7:1957-1962(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;
RA Hibbs M.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
receptor tyrosine kinase.";
RL Genes Dev. 8:399-413(1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9132866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morée B.;
RT "Comparison of EGF receptor sequences as a guide to study the ligand
binding site.";
RL Oncogene 6:673-676(1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eisenger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
```

tyrosine phosphate.

-!- SUBCELLULAR LOCATION: Type I membrane protein.

-!- MISCELLANEOUS: Binding of EGF to the receptor leads to dimerization, internalization of the EGF-receptor complex, induction of the tyrosine kinase activity, stimulation of cell DNA synthesis, and cell proliferation.

-!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X79887; CAA55587.1; -

EMBL; U03425; AAA17899.1; -

EMBL; X59698; CAA42219.1; -

EMBL; L06864; AAA53029.1; -

EMBL; Z12608; CAA78249.1; -

HSSP; P11362; 1FGK.

MGI; MGI:95294; Egfr.

InterPro; IPR000494; EGFR_L domain.

InterPro; IPR000719; Euk_pkinase.

InterPro; IPR002174; Furin-like.

InterPro; IPR001245; Tyr_pkinase.

Pfam; PF00069; pkinase; 1.

Pfam; PF00757; Furin-like; 1.

Pfam; PF01030; Recep_L domain; 2.

ProDom; PD000001; Euk_pkinase; 1.

SMART; SM00261; FU; 3.

SMART; SM00219; TyrKc; 1.

PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.

Transmembrane; Glycoprotein; Receptor; Signal; Transferase;

Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.

SIGNAL 1 24 POTENTIAL.

CHAIN 25 1210 EPIDERMAL GROWTH FACTOR RECEPTOR.

DOMAIN 25 647 EXTRACELLULAR (POTENTIAL).

DOMAIN 671 1210 POTENTIAL.

DOMAIN 671 1210 CYTOPLASMIC (POTENTIAL).

REPEAT 75 300 APPROXIMATE.

REPEAT 390 600 APPROXIMATE.

DOMAIN 1028 1071 SER-RICH.

DOMAIN 714 981 PROTEIN KINASE.

NP_BIND 720 728 ATP (BY SIMILARITY).

BINDING 747 747 ATP (BY SIMILARITY).

ACT_SITE 839 839 BY SIMILARITY.

DISULFID 130 199 BY SIMILARITY.

DISULFID 194 207 BY SIMILARITY.

DISULFID 215 223 BY SIMILARITY.

DISULFID 219 231 BY SIMILARITY.

DISULFID 232 240 BY SIMILARITY.

DISULFID 236 248 BY SIMILARITY.

DISULFID 251 260 BY SIMILARITY.

DISULFID 264 291 BY SIMILARITY.

DISULFID 295 307 BY SIMILARITY.

DISULFID 311 326 BY SIMILARITY.

DISULFID 329 333 BY SIMILARITY.

DISULFID 506 515 BY SIMILARITY.

DISULFID 510 523 BY SIMILARITY.

DISULFID 526 535 BY SIMILARITY.

DISULFID 539 555 BY SIMILARITY.

DISULFID 558 571 BY SIMILARITY.

DISULFID 562 579 BY SIMILARITY.

DISULFID 582 591 BY SIMILARITY.

DISULFID 595 617 BY SIMILARITY.

DISULFID 620 628 BY SIMILARITY.

DISULFID 624 636 BY SIMILARITY.

MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).

MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

[illegible]


```
Db      -536  CNLYDGEFREFENGSIVCEDPOCEKQWEDGLLTCHGPGPDNCTKCSHFDPNCVKECPD 595
Qy      603  GUKPDLSYMPIWKFPDEBEGACQPCPNCTHSCVDLDDKGC-----PAEQRASPL 651
Db      596  GLQGANSF--IFKYADPDRECHPCPNCTQGCNGPTSHDCIYYPMTGCHSTLPOHAR-TPL 652
Qy      652  TSIVSAVV--GILLVVLGVVFGILIKRROOKIRKYTMRRLLLOTELVELPLTSGAMPNQA 710
Db      653  --IAGVIGGLFILVIVGUTFAVYVRKSIK-KRALRRFL-ETELVEPLTSGTAPNQA 708
Qy      711  QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENYKIPVAIKVLRENTSPKANKEILDE 770
Db      709  QLRILKETELRKVKVLGSGAFGTVYKGIWIPDGENYKIPVAIKVLRENTSPKANKEILDE 768
Qy      771  AYVMAGVSPYSRLLIGLITSTVQLVTLQMPYGCLLLDHVRNRLGSLQDILLNWCMTA 830
Db      769  ALIMASMDPHLVRLLGVLCSPTIQLVTLQMPHGCLEYYVHEHKNIGSQLLNWCMTA 828
Qy      831  KGMSVLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDDDEYHADGGKVPKWA 890
Db      829  KGMVLEERLLVHRDLAARNVLKSPNHVKITDFGLARLLDEGEKYNADGKMPKWA 888
Qy      891  LESILRRRTHOSDVMYSYGVTWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950
Db      889  LECIHYRKFTHOSDVMYSYGVTWELMTFGKPYDGIPAREIPDLLEKGERLPQPPICTID 948
Qy      951  VTMVWKCMIDSECRPRFELVSRMRARDPQRFVWIONED-LGPASPLSTFTVRSLL 1009
Db      949  VTMVWKCMIDSECRPRFELVSRMRARDPQRFVWIONED-LGPASPLSTFTVRSLL 1008
Qy      1010  EDDDDGDLVDAEYLVPQGFCCPDAPCAGAGVHRRHSSSTRSGGDLTLGLEPSEEE 1069
Db      1009  DEEDLEMDDAEYLVP-QAFNIPP-----IYTSRAIDNRS-----EIGHSPPPAY 1056
Qy      1070  APRS-----PLAP-SEGAGSDVDFDGLMGAAKGLQS 1100
Db      1057  TMSGNQFYVRDGFAGAEQGVSVYRAPTSITPEAPVAGATAEIFFDDSCCNGTLRKPA 1116
Qy      1101  LPTHBPSPLORSEDTVPPLPS-----ETDGYVAPLTCSPQPEYVQPDVVRPQPPSR 1153
Db      1117  PHVQDSSTQRYSDPTVFAPERSPRGELDEEGYMTMRDKPKQEYLNPE-----1167
Qy      1154  EGPLPAARPAAGATLERAKTLPQKGVKVDVAFAGGAVENPEYLPQGGAAAPQHPHPPA- 1212
Db      1168  ENPFVSR-----KNGDLQ-----ALDNPETHNASNG-----PPKAE 1199
Qy      1213  -----FSPAFDNLVYWDQDPPPERGA--PPSTF 1237
Db      1200  DEYVNEPLYNTFANTLGRKAEYLNKILSMPEKAKKAFDNDYWNHSLPPTSLQHPDYL 1259
Qy      1238  KGTPT-----AENPEYL 1249
Db      1260  QEYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 7
ERR4_RAT
ID  ERB4_RAT  STANDARD;  PRT;  1308 AA.
AC  Q62956; Q922N7;
DT  15-DEC-1998 (Rel. 37, Created)
DT  15-JUN-2002 (Rel. 41, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN  ERB4 OR TYRO-2.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]_TaxID=10116;
RP  SEQUENCE FROM N.A.
RC  TISSUE=Heart;
RX  MEDLINE=98221155; PubMed=9553078;
RA  Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
```

```
RA  Marchionni M.A., Kelly R.A.;
RT  "Neuregulins promote survival and growth of cardiac myocytes.
RT  Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT  ventricular myocytes.";
RL  J. Biol. Chem. 273:10261-10269(1998).
RN  [2]
RP  SEQUENCE OF 848-901 FROM N.A.
RC  TISSUE=Sciatic nerve;
RX  MEDLINE=91222560; PubMed=2025425;
RA  Lai C., Lemke G.;
RT  "An extended family of protein-tyrosine kinase genes differentially
RT  expressed in the vertebrate nervous system.";
RL  Neuron 6:691-704(1991).
RN  [3]
RP  SEQUENCE OF 1031-1198 FROM N.A.
RC  STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX  MEDLINE=97184212; PubMed=9030624;
RA  Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT  "Expression of neuregulins and their putative receptors, ErbB2 and
RT  ErbB3, is induced during Wallerian degeneration.";
RL  J. Neurosci. 17:1642-1659(1997).
CC  -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC  2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC  NTK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC  NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC  -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC  tyrosine phosphate.
CC  -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC  RECEPTORS (POTENTIAL).
CC  -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC  -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC  NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC  OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC  RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC  HEART.
CC  -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC  RESIDUES (BY SIMILARITY).
CC  -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
EMBL; AF041838; RAD08899.1; --
EMBL; U52531; AAC53051.1; --
HSSP; P11362; 1FGK.
InterPro; IPR000494; EGFR_L domain.
InterPro; IPR000719; Euk_pkinase.
InterPro; IPR002174; Furin-like.
InterPro; IPR001245; Tyr_pkinase.
InterPro; IPR004019; YLP motif.
Pfam; PF00757; Furin-like; 1.
Pfam; PF00069; pkinase; 1.
Pfam; PF01030; Recep_L domain; 2.
Pfam; PF02757; VLP; 2.
PRINTS; PR00109; TYRKINASE.
ProDom; PD000001; Euk_pkinase; 1.
SMART; SM00261; FU; 4.
SMART; SM00219; TyrKc; 1.
PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25
FT CHAIN 26 1308
FT DOMAIN 26 651
FT TRANSMEM 652 675
FT DOMAIN 676 1308
CYTOPLASMIC (POTENTIAL).
```



```
Db 590 IPRCPHGLGDDTL-TWKYADKMGQCPCHQNCQTQCGSLGCRGD-IVSHSLAVG 647
Qy 657 AVUGILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRLK 716
Db 648 LVSGLLITVALLIVLLRRRRIK-RKRTIRCLQELVEPLTPSQAPNQAFRLIK 706
Qy 717 ETELKRVKVLGSGAGFYKGIWIPDGENKIPVAIKVLRNTPSKANKEILD EAYVNA 776
Db 707 ETEFKDRLVLSGAGFYKGLMNPGENIRIPVAIKVLRNTPSKANKEILD EAYVNA 766
Qy 777 VGSFVYVRLGICLTSTVLTQMPYCGCLLDHVRNRLGSLQDLINWCQIAKGSYL 836
Db 767 VDRPHVCRLLGICLTSAVLQVLTQMPYCGCLLDYVQRQERICGQWLLNWCQIAKGMNYL 826
Qy 837 EDVRLVHRDLAARNVLKPNHVKITDFGLARLLDIDETEHADGKVPKIKMALESLR 896
Db 827 EERHLVHRDLAARNVLKPNHVKITDFGLSKLLTADKEYOAGGKVPKIKMALESILQ 886
Qy 897 RRFTHQSDVWSYGVTVVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYIMV 956
Db 887 WTYTHQSDVWSYGVTVVWELMTFGSKPYDGIPIAREIPDLLEKGERLPPOPICTIEVYMIIL 946
Qy 957 KCMWIDSECPREFELVSESRWARDQRFVVIQNEIDLGPASPLDSTFYRSLLED DMDG 1016
Db 947 KCMWIDSPRRPRELVEGFSQWARDPSRYLVIOG---NLPSLSDRLFSRLSSDD--D 1001
Qy 1017 LVDAEYLVPOOGFCFDPAPGAGGMVHRRHSSTRSGGDLTLGLEPSEEAAPRPLA 1076
Db 1002 VVDAEYLLPKRI-----NRQGS-----EPCI 1024
Qy 1077 PSEGAGSDVFDGLGMAAGLQSLPDPSPLOYSEDPTV-PLPSETDGYVAPLTCSP 1135
Db 1025 PPTGH-----PVRENSITLRNISDPTQNALEKDLGDH----- 1056
Qy 1136 QPEVVPDPVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTLSPKNGVVK 1182
Db 1057 --EVNPGGETSRRLSDIYNPTVEDTLDGMPVLSLSQEAETNFSPEYLTNTQNSL-- 1112
Qy 1183 DVFAFGAVENPEYLTQGGAAPOPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPT 1242
Db 1113 -PLVSSGSMDDPDY---QAG-----YQAFAF-----LPQTGALTNGMGFLPA 1149
Qy 1243 AENPEYLG 1250
Db 1150 AENLEYLG 1157

RESULT 9
ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors."
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
```

```
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene."
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RX TISSUE=Placenta;
RX MEDLINE=9328282; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase."
RL Blochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -!- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M29366; AAA35790.1; --
CC EMBL; M34309; AAA35979.1; --
CC EMBL; S61953; AAB26935.1; --
CC PIR; A36223; A36223.
CC HSSP; P11362; IFGK.
CC Genew; HGNC:3431; ERBB3.
CC MIM; 190151.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; FALSE NEG.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
CC PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202
```

FT	DISULFID	210	BY SIMILARITY.
FT	DISULFID	214	BY SIMILARITY.
FT	DISULFID	226	BY SIMILARITY.
FT	DISULFID	227	BY SIMILARITY.
FT	DISULFID	231	BY SIMILARITY.
FT	DISULFID	243	BY SIMILARITY.
FT	DISULFID	246	BY SIMILARITY.
FT	DISULFID	255	BY SIMILARITY.
FT	DISULFID	259	BY SIMILARITY.
FT	DISULFID	290	BY SIMILARITY.
FT	DISULFID	301	BY SIMILARITY.
FT	DISULFID	305	BY SIMILARITY.
FT	DISULFID	323	BY SIMILARITY.
FT	DISULFID	327	BY SIMILARITY.
FT	DISULFID	500	BY SIMILARITY.
FT	DISULFID	509	BY SIMILARITY.
FT	DISULFID	517	BY SIMILARITY.
FT	DISULFID	520	BY SIMILARITY.
FT	DISULFID	529	BY SIMILARITY.
FT	DISULFID	533	BY SIMILARITY.
FT	DISULFID	549	BY SIMILARITY.
FT	DISULFID	552	BY SIMILARITY.
FT	DISULFID	556	BY SIMILARITY.
FT	DISULFID	573	BY SIMILARITY.
FT	DISULFID	576	BY SIMILARITY.
FT	DISULFID	589	BY SIMILARITY.
FT	DISULFID	610	BY SIMILARITY.
FT	DISULFID	613	BY SIMILARITY.
FT	DISULFID	621	BY SIMILARITY.
FT	DISULFID	629	BY SIMILARITY.
FT	CARBOHYD	126	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	250	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	250	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	353	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	353	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	408	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	414	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	437	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	437	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	469	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	522	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	522	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	566	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	616	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	616	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPLIC	141	EILSGVYIEKNDLCHMDTIDWRDVIIVRDRAEIVVKNR
FT			SC -> GQFMPWPGTLPQADQVLLDDPPRLTLUSASK
FT			VPTLAIV (IN SHORT ISOFORM).
FT	VARSPLIC	184	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	560	E -> G (IN REF. 2).
FT	CONFLICT	1064	E -> G (IN REF. 2).
SQ	SEQUENCE	1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;	
Query Match			
Best Local Similarity 40.5%; Pred. No. 2.2e-121;			
Matches 532; Conservative 195; Mismatches 452; Indels 135; Gaps 34;			
QY	10	GLLLALLPGAA--STQVCTGTDMLRLPASPTHLDMLRLHYQGGVQVQGNLELYLPT 67	
DB	11	GLLFLARGSEVGNQAVCPGLTNGLSVTDGAENQYTLKLYERCEVYMGNEIVLIGH 70	
QY	68	NASLFLQDIQEVQVGLIAHNOVQVPLQRLIRIVRGTLQFEDNVALAVLDNGDPLNNTT 127	
DB	71	NADLSFLQWIREVTGVVLVAMNEFTLPLNLRVVRGTQVYDQKFAIFVW-----LYNYT 125	
QY	128	PVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFPKNNQALTLIDTN 187	
DB	126	---NSSHALRQLRLTLTEILSGGVYIEKNDLCHMDTIDWRDVIIVRDRAEIVVKN 178	
QY	188	RSRACHPCSPMCKSRGWESSEDOSLRTVCAGGC-ARCKGPLPTDCHQCAAGCTG 246	
DB	179	NGRSCPPCHEVCCKG-RCMVPGSEDCQTLTKTICAPQCNHGCFCGPNPNOQCCHDECAGGCSG 237	
QY	247	PKHSCLACLHFNHSGICELCPALVYNTDTFESMPNPEGRYTFGCAVCVACPYNLYST 306	
DB	238	PQDTCFACRHENDSGACVPRCPQPLVYNKLTFOLEPNHTKYQYGGVCVASCNPNFV-V 296	
QY	307	DVGSCTVLCPHLNQEVTAEDGTORCEKCKSPCARVCYGLGMOYI KANSKF--IGITELE- 363	
DB	297	DQTSVRACPDPMKMEVD-KNGLKMGCEPCGGLCPKACEGTG-----SGSRFQTVDSNIDG 350	
QY	364	FAGCKKIFGSLAFLESFDGDPASNTAPLOPQLOVFTELEITGYLISAWPDSLPLDS 423	
DB	351	FVNCTKIIGNLDFLITGLNGDPMHKIPALPDPEKJNVFRTVREITGYLNIQSWPPHMFNS 410	
QY	424	VFQNLQVIRGRILHNGAYS-LTLOGLIGISWLGRSLRELGSGLALIHNTHLCFVHTVPW 482	
DB	411	VFSNLTITIGRSLYNRGFSLLIMKMLNVTSLGFRSLKEISAGRIYISANRQCLCYHHSLNW 470	

QY	483	DQLFRNPHOALLHTA-NRPEDEBCVGEGLACHOLCARGHCKWPGPTQCVCNSQFLRGQCV 541	
DB	471	TKVLRGPTTEERLDIKHNRPRRDCVAEGKVCDCPLCSSGGCWCWPGQCLSCRNYSRGVCV 530	
QY	542	ECRVLQGLPREYVNAHCLCPHCEQOPQNGSVTCFGEADQCVACAHYKDPFCVACP 601	
DB	531	THCNFLNGEPREFAHEAEFCFCHPECPMEGTATCNGSGSDTCAQCAHFRDGHCVSSCP 590	
QY	602	SGVKPDLSPYMPYKPFDEEGACQPCPINCTHSCVDLDDKGCAPAEORA---SPLTSTVSA 657	
DB	591	HGVLG--AKGPIYKYPDVQNECRPCHENTCGCKGPELQCLQTLVLICKTHLTMAITV 648	
QY	658	VVGILLVVLGVVFGILIKRQOKIR-KYTMRLLOTELVELPELTSGAMPNOAQMRLK 716	
DB	649	IAG--LVWIFMMLGGTFLYWRGRRIQNKRAMRYLERGESIEPLDPS-EKANKVLARIFK 705	
QY	717	ETELRKVKVLGSGAGCTGVYKGIWIPGENVKIPVAIKVLRNTSPRANKELIDEAYVMAG 776	
DB	706	ETELRKLVKLVGSGVGTGTVHKGVMPGESIKVIEDKSGROSFOAVTDHMLAIGS 765	
QY	777	VGPSVYSRLGICLTSTVOLTPQMPYGLLDHVRNRRGLSGQDLLNMCQIAKMSYL 836	
DB	766	LDAHIVRLVLLGLCPGSSLOLVQYLPGLSLLDHVRQHRGALGQDLLNHWQVQIAKMYL 825	
QY	837	EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYHADGKVPKIMMALESILR 896	
DB	826	EEHGMVHRNLAAARNVLLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTPIKMALESIFH 885	
QY	897	RRTHQSDVWSGVTVWELMTGAKYDGPAREIPOLLEKGERLQPPCTIDVYVMV 956	
DB	886	GKYYHQSDVWSGVTVWELMTGAEYAGRLAEVDPDLLEKGERLAQPPCTIDVYVMV 945	
QY	957	KCWMIDSECRPRELVSFESRMARDPQRFVIVIONEDLQPA---SPLDSTFYRSLLEDD 1013	
DB	946	KCWMIDENIRPTFKELANEFTRMARDPPRYLVIKRES-GPGIAGPEPHGLTNNKKEVE 1004	
QY	1014	MGDLVDAEYLVPPQGFCCPDPAAGAGMVHRRSSSTRSGGDLTLGLEP-SEEBAPR 1072	
DB	1005	LEPELDLDLEAEED-----NLATTLSALSPLVGLTNNRRPGSQ 1045	
QY	1073	SPLASREGAGSDVFDGLMGAAKGLQSLPTH-DPSFLORYSEDTVPPLD-----SETD 1125	
DB	1046	SLLSPSSGY-MPMNQNLGSCQESAVSGSSERCPVSLH-----PMRPGCLASSE 1098	
QY	1126	GVVA-----PLTCSQPE-----YVNPQDVVRPOPSPRECP----- 1156	
DB	1099	GHVTGSEAELOEKVSMCRSRSRSPRGRDSAYHSQRHSLLTPTVPLSPGLEEDVNG 1158	
QY	1157	--LPAARPAAGATLEAKTLSP-GKNGV-----KDVFAFGGAVENPEYLTPOGGAAPQ 1207	
DB	1159	YVMPDTHLKGTPSSREGTILSSVGLSVLGTEDDED-----EYEVYNNRRRRHSP-P 1209	
QY	1208	HPPAPSPAFDNLVYWD-----QDPPERGAPPSTFKGTPTAENPEYL 1249	
DB	1210	HPPRPSLEELGYEYMDVGDLSASLGSTQSCPLHPVIMPTAGTTTDEDEYEM 1263	
RESULT 10			
ID	ERB3_RAT	STANDARD;	PRT; 1339 AA.
AC	Q62799; Q62955;		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)		
DE	(c-erbB3).		
GN	ERB3.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Kolland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
recombinant protein.";
RL Gene 165:279-284(1995).
RN [2]
RP REVISIONS TO 85; 513 AND 565.
RA Hellyer N.J., Kolland J.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuroligins and their putative receptors, ErbB2 and
ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTKA.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
(POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
PHOSPHATIDYLINOSITOL 3-KINASE.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

CC EMBL; U29339; AAC28498.2; --
DR EMBL; U52530; AAC53050.1; --
DR HSSP; P11362; LFQK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19
FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 644 662 POTENTIAL.
FT DOMAIN 663 1339 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 183 259 CYS-RICH.
FT NP_BIND 707 964 PROTEIN_KINASE.
FT NP_BIND 713 721 ATP (BY SIMILARITY).
FT BINDING 740 740 ATP (BY SIMILARITY).
FT ACT_SITE 832 832 BY SIMILARITY.
FT DISULFID 186 194 BY SIMILARITY.
FT DISULFID 190 202 BY SIMILARITY.
FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.

FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1028 1028 L -> P (IN REF. 3).
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBDFD1E CRC64;

Query Match 34.5%; Score 2351.5; DB 1; Length 1339;
Best Local Similarity 40.9%; Pred. No. 1.4e-117;
Matches 525; Conservative 171; Mismatches 430; Indels 159; Gaps 35;

QY 3 LAALCRWGLLLALLPPGAA---STQVCTCTDKMLRLPASPTHLDMLRLHYQSCVQVQGN 59
DB 7 LQVLC---FLSLARGSEMGSQAVCPGTNLGLSVTGADNOYQTYLYKDYKEVVMGN 62
QY 60 LELTYLPTNASLFLQDIOEVQGYVLIAHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDN 119
DB 63 LEIVLTGHNADLSFLQWIREVTGYVLVAMNEFSLPLPNLRVVRGTVQVYDGKFAIFVM-- 120
QY 120 GDP LNTTPTVTGASPGGLRELQRLSLTEILKGVLIOQNPOLCYQDTILWKDILFHKNQL 179
DB 121 ---LNYNT---NSSHALRQLKFTQLTEILSGGVYIEKNDKLMOTIDWRDVRVR-- 170
QY 180 ALTLTDNRSRACHPCSPMKSGESSEDCOSLTRVTVCAGGC--ARKCKGLPTDCCHE 238
DB 171 GAIVVKNNGANCPCHVECKG--RCWGPDPDCQILTKTICAPQCNRCFCGPNQCCHD 229
QY 239 QCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
DB 230 ECAGGCGSPQDTCFACRFRFNDSGACVPRCPPLVYNKLTFFOLEPNPHTKYQYGGVCVAS 289
QY 299 CPYNYLSTDVGSCTLVCP LHNQEVTAEDGTQCEKCKSPCARVCYGL--GMQYIKANSKF 356
DB 290 CPHNFV--VDQTFVCRACPPDKMEVD--KHGLKMCPCGGLCPKACEGTSGSRVQTVDDSN 347
QY 357 IGITELEFAGCKKIFGSLAFLESPDGDPSANTAPLOPELOVFEETLEITGVLYISAWP 416
DB 348 ID----GFVNCIKILGNLDFLTGLNVDPWHKIPALDPEKINVFRVREITGYLNTQSWP 403
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYS--LTLOGLGISWGLRSLRSLRELGSLALHNNTHLC 475
DB 404 PHMNFVSFSLNTTIGRSLYNGFSLIMKLNVTSLGFRSLKEISAGRYVISAQNLQC 463
QY 476 FVHTVPWDLFRNPHQALLHTA--NRPEDEVCEGLACHOLCARGHCHWGFGPTQVCNCSQF 534
DB 464 YHHSNLWTRLLRGPSEERLDIKYDRPLGCELAGKVCYDPLCSGGGCGWPGGQCLSCRNY 523
QY 535 LRQGEVCEBCRYLQGLPREYVNNARHCLPCHPCQPQNGSVTCFPGPADOCVACAHYKDP 594
DB 524 SREGVCVTHCNFLQGEPRFVHEAQCFCSPCLPMEGTSTNGSGSDACARCAHFRDGP 583
QY 595 FCVARCPGKPDLSYMPIWKFPDBEGACQPCPINCTHSC--VDLDDKCPAEQASPLT 652


```
Db      584 HCVNSCPHILG--AKGPIYKTPDAQNECRPCHENTQCNCPELODCQLGQAEVLMSRPH 641
QY      653 SIVSAVGGILLVVVLGVFGILIKRQOKIR-KYTMRRLLQLOTELVEPLTPSGAMPNQAO 711
Db      642 LVIAVTVG--LAVILMILGSGFLYWRGRIQNKRAMRRYLERGESIEPLDPS-ERAKVYL 698
QY      712 MRILKETLKVVLGSGAFGVYGIWIIPDGENVKIPVAIKVLRENTSPKANKILDEA 771
Db      699 ARIFKETLRLKVLGSGVFGVHRGIWIPEGESIKIPVCIKVIEDKSGRQSFQAVTDHM 758
QY      772 YVMAGVGSPPYSRLLIGLCTSTVOLVTLQMPYGCLLDHVRENRLGSLDGLNMCWLAK 831
Db      759 LAVGSLDHAHVIRLLGLCPGSSQLVTVQLPGLSLLDHVHQKRETLGPPOLLNMGVQLAK 818
QY      832 GMSYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETRYHADGGKVPKMMAL 891
Db      819 GMYLLEHSMVHRDLARNVLMKSPQVQVADFGVADLLPPDDKQLLHSEAKTPIKMMAL 878
QY      892 ESILRRRTHQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLEKGERLPPOPTCTIDV 951
Db      879 ESIHFGKTHQSDVMSYGVTVWELMTFGAEPYAGRLABIPDLEKGERLAQOQICTIDV 938
QY      952 YIMVWKMWIDSECRPRFRELSEFSRMARDPQRFVWIONEDLGPASPLDSTFYRSLLIED 1011
Db      939 YVMVWKMWIDENIRPTKELANEFTRMARDPRVLYIKRAS-GRGTP--PAAEFSLVT 995
QY      1012 DMGDLVDAEYVLVPOQFFCPDPAPAGGMVHRRHSSTRSGGDLTLGLEPSEE--- 1068
Db      996 KEL-----QEALEPEL-----DLDLDLEAEEGLA 1021
QY      1069 -----EAPRSLAPSEG-----AGSDVFDGDLGMAKGLQSLPT 1103
Db      1022 TSLGSAISLPTGTLTRPGSQSLSPSSGYMPMNOSSLGAECLDSAVLGGREFQSRPISL 1081
QY      1104 HDPSLPQRYSEDTPLPSETDGVV---APL-----TC-----SPOPE---YVNOP 1143
Db      1082 H-PIPRGR-----PASESEGHVTGSAELQEKVCSRSRSPRPRGDSAYHSQR 1133
QY      1144 DVPPOPPSPREGP-----LPAARPAGATLERAKTSLP-KGNGVV-----KDVF 1185
Db      1134 HSLTLPTVPLSPGLEEDNGCYVMPDTHLRGASSSREGTLLSGVLSVLGTEEDED-- 1191
QY      1186 AFGGAVENPEYLTPOGGAPQHPH 1210
Db      1192 -----EEYEMNRKRGSP-PRPP 1209

RESULT 11
ID      EGFR_DROME      STANDARD;      PRT: 1426 AA.
AC      P04412; O61601; Q9W2G0; P81868;
DT      13-AUG-1987 (Rel. 05, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE      (Gurken receptor) (torpedo protein) (Drosophila relative of ERBB).
GN      EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC      Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC      Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
RX      NCBI_TaxID=7227;
[1]
RP      SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX      MEDLINE=94350209; PubMed=8070664;
RA      Clifford R., Schubach T.;
RT      "Molecular analysis of the Drosophila EGF receptor homolog reveals
RL      that several genetically defined classes of alleles cluster in
RN      subdomains of the receptor protein."
[2]
RP      Genetics 137:531-550(1994).
```

```
RA      Clifford R., Schubach T.;
RL      Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
[3]
RP      SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX      MEDLINE=85124611; PubMed=2982499;
RA      Livneh E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;
RT      "The Drosophila EGF receptor gene homolog: conservation of both
RL      hormone binding and kinase domains."
[4]
RP      Cell 40:599-607(1985).
RP      SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX      STRAIN=Oregon-R; Tissue=Embryo.
RX      MEDLINE=87002474; PubMed=3093080;
RA      Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT      "Alternative 5' exons and tissue-specific expression of the
RL      Drosophila EGF receptor homolog transcripts."
[5]
RP      Cell 46:1091-1101(1986).
RP      SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
ANALYSIS.
RX      MEDLINE=99102120; PubMed=9882502;
RA      Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT      "Several levels of EGF receptor signaling during photoreceptor
RL      specification in wild-type, Ellipse, and null mutant Drosophila."
[6]
RP      Dev. Biol. 205:129-144(1999).
RP      SEQUENCE FROM N.A. (ISOFORM TYPE I).
RX      STRAIN=Berkelley;
RX      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Calniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA      Ballew K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA      Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA      Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA      Fosler K.C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA      Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA      Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kinsella C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA      Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA      Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA      Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA      Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA      Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA      Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT      "The genome sequence of Drosophila melanogaster."
[7]
RP      Science 287:2185-2195(2000).
RP      SEQUENCE OF 959-1078 FROM N.A.
RX      STRAIN=Daekwanryeong;
RX      MEDLINE=85137938; PubMed=2983232;
RA      Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT      "A Drosophila genomic sequence with homology to human epidermal
RP      REVISIONS.
```

RT growth factor receptor.";
 RN Nature 314:178-180(1985).
 RN [8]
 RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
 RP ANALYSIS.
 RX MEDLINE=92038942; PubMed=1936959;
 RA Raz E., Schejter E.D., Shilo B.Z.;
 RT "interallelic complementation among DER/fib alleles: implications for
 RT the mechanism of signal transduction by receptor-tyrosine kinases";
 RL Genetics 129:191-201(1991).
 RN [9]
 RP REVIEW.
 RX MEDLINE=97248481; PubMed=9094709;
 RA Pertimon N., Perkins L.A.;
 RT "There must be 50 ways to rule the signal: the case of the Drosophila
 RT EGF receptor";
 RL Cell 89:13-16(1997).
 CC -!- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OOCYTE. IN THE EMBRIO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOEROSA
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS: TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THORACIC AND ABDOMINAL GANGLIA.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF052754; AAC08536.1;
 DR EMBL; AF052753; AAC08536.1; JOINED.
 DR EMBL; AF052754; AAC08535.1;
 DR EMBL; AF052752; AAC08535.1; JOINED.
 DR EMBL; K03054; AAA51462.1;
 DR EMBL; K03417; AAA51460.1;
 DR EMBL; K03416; AAA50965.1;
 DR EMBL; K03418; AAA51461.1;
 DR EMBL; AF109077; AAD26134.1;
 DR EMBL; AF109078; AAD26132.1;
 DR EMBL; AF109082; AAD26132.1; JOINED.
 DR EMBL; AF109078; AAD26133.1;
 DR EMBL; AF109084; AAD26133.1; JOINED.
 DR EMBL; AF109079; AAD26130.1;
 DR EMBL; AF109081; AAD26130.1; JOINED.
 DR EMBL; AF109079; AAD26131.1;
 DR EMBL; AF109083; AAD26131.1; JOINED.
 DR EMBL; AF109080; AAD26135.1;
 DR EMBL; AE003454; AAF46732.1;
 DR EMBL; X02293; CAA26157.1;
 DR EMBL; X78920; CAA55523.1;
 DR EMBL; X78918; CAA55521.1; JOINED.

DR EMBL; X78919; CAA55522.1; JOINED.
 DR PIR; A00640; GQFPE.
 DR HSSP; P11362; IFCK.
 DR FlyBase; FBgn0003731; Egfr.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk pkinase; 1.
 DR SMART; SM00261; EU; 7.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
 KW Developmental protein.
 FT SIGNAL 1 30
 FT CHAIN 31 1426
 FT DOMAIN 31 868
 FT TRANSMEM 869 889
 FT DOMAIN 890 1426
 FT DOMAIN 938 1198
 FT NP_BIND 944 952
 FT BINDING 971 971
 FT ACT_SITE 1063 1063
 FT MOD_RES 902 902
 Query Match 28.6%; Score 1952; DB 1; Length 1426;
 Best Local Similarity 32.7%; Pred. No. 2.5e-96;
 Matches 468; Conservative 183; Mismatches 429; Indels 350; Gaps 42;
 QY 24 QVCTGDMKRLPASPEHDLMLRHLVQCQVQGNLELTLYPT-NASLSFLQDIOEVOG 82
 DB 100 KICIGTKSRLSPVSNKEHHYRLRDYNTCTVVDGNLKLTLWLPNENLDSFLDNIREVTG 159
 QY 83 YVLIHNOVQVPLQRLRIVRGTLF-----EDNYALAVLDNGDPLNNTPTVTGASPGGL 137
 DB 160 YILISHVDVKKVFPKQLQIRGTLFSLSVSEKVALFV-----TYSKM 203
 QY 138 RELQLRSLTEILKGVVLQORNPOLCYQDTILWKDIFHKNNQLALTLIDNRSRACHPCSP 197
 DB 204 YLEIPDLRDVLNGQVGFHNNYNLCHMRTIQWSEIVSNGTDAYYNYDFAPERECPKCHE 263
 QY 198 MCKGSRGSGESDCCQSLTRITVCAGCA--RCKGPLPTDCCHQCQAAGCTGPKHSDCLAC 255
 DB 264 SETHG-CHGEGPKNCQKSKLTCSPQACGRCYGPKPRECHLFCAGGCTGTQKDCIAC 322
 QY 256 LFNHSGICELHCPALVTYNTDTTFESMPNPEGRYTFGASCVTACPNYVLTSDVGSCTLVC 315
 DB 323 KNFFDEAVSKCECPPMKYNPTTYVLETNPEKAYGATCVKECP-GHLLRDNAGACVRSC 381
 QY 316 PLHNOEVTAEDGTQRCCKSKPCARVCYGLGHQYKANSKFGITEL-----EPAGCKK 369
 DB 382 PQDKMDKGE-----CVPCNGPCPKTC-----PGVTVLHAGNIDSPFNCTV 422
 QY 370 IFGSLAFLPESFDG--DPASNTA-----PLQEOLOVFTLEETIGYLYISAWPDSLDP 421
 DB 423 IDGNIRILDQTFSGFDQVYANYTGMCPRIPLDPERREVFSTVKEITGYLVNIGTHPOFRN 482
 QY 422 LSVFONQVIRGRILHNGAY-SLTLOGLIGISWLGRLSRLSGLSLALIHNNHLCFVHTV 480
 DB 483 LSYFRNLETIHGRQLMSEMFALAIVKSSLSLEMRNLKQISSGSVWIQHNRDLVCVSN 542
 QY 481 PWDLFRNPHQALLHTANRPEDECYVGEGLACHOLCARGHCWGPFGTQCVCNCSQFLRGQBC 540
 DB 543 RMPAIQKEPEQKVWVNNENLRADLCENKGTICSDQCNEBGCWAGTDOCLTCNFFNNGTC 602
 QY 541 VEECRVLQGLPREYVYNARHCLPCHPECPQNGSVTCFCGPEADQCACAHYKDPFCVARC 600

```

Db 603 IADCGYISNAYK--FDNRFTCKIHPCCR-----TCNGAGADHCQECVHVRDQGHQCVSBC 654
Qy 601 P-----SGVK-----PDL 608
Db 655 PKNKYNDRGVRECHATCGDCTGPKDTIGIGACTTCNLA1INNDATVVKRCLLKDDKCPD- 713
Qy 609 SYMPTWKF--PDEEGACQP-----CPI-----NCTH-----632
Db 714 GY--FWEYVHPQEGSLKPLAGRAVCRKCHPLCELCTNYGYHEQVCSKCTHYKRRQCBET 771
Qy 633 -----SC-----VDLDKG-----641
Db 772 ECPADHYTDEQRECFORHPECNGCTGPGADDCKSCRNFKLFDANETGYPVNSTWPNCTIS 831
Qy 642 -CPAQR-----ASPLTS-----IVSAVVGILLVVVLGVVFGI 673
Db 832 KCPLEMRHNYOVYTAIGPYCAASPPRSSKITANLDVNMFIITGAVLVPTICILCV--T 889
Qy 674 LKRRQOKIRKYT--MRRLLOETELVEPLTPSGMNPQAMQMRILAKETELRKVKVLGSGAF 731
Db 890 YICROKQAKKETVKNMTALSCEDSEPLRPSNIGANLCKLRIVKDAELRKGVGLMGAP 949
Qy 732 GTVYAGIWIPIGDNENKVIPIVAIKVLRNTPSKANKEILDEAYVMAGVGSFVYGRLLIGICLT 791
Db 950 GRVYGVVYPEGENVKIPVAIKELLKSTGAESSEEFLEAYIMASEHVNLLKLLAVCMS 1009
Qy 792 STVQLVLTQMPYGCCLLDHVRNRRGLSQDLLNWCQIAKMSYLEDVRLVHRDLAARNV 851
Db 1010 SQMMLITLQPLGLLDVYVNRNRDKTIGSKALLNWSIOIAKMSYLEEKRLVHRDLAARNV 1069
Qy 852 LVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRFTHQSDVWSYGVY 911
Db 1070 LVQTPSLVKITDFGLAKLLSSDNEYKAAAGGKMPKIMLALECIKRNKRVFTSKSDVWAFGYT 1129
Qy 912 VNELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYIMVWKWMIDSECRPRFE 971
Db 1130 IWELTTFGQPHENIPAKDIPDLIEVLGKLEQPEICSLDIYCTLLSCHWLDAMRTFKQ 1189
Qy 972 LVSEFSRMARDPQRFVVIQNEDLG--PASPLDSTFYRSLLED--DMGDLDVAEYLYP 1026
Db 1190 LTTVFAEFAFGVLAIPGDKFTRLPA-----YTSQDEKDLIRKLAPTTDGSIAIAK 1242
Qy 1027 QQGFPCPDAPGAGGVHHRSSSTRSGGDLTLGLEPSEAP-----RSLAPSEG 1080
Db 1243 PDYIQPKAAGPS-----HRTDCT-----DEMPKLNRYCKDPSKNKSS 1281
Qy 1081 AGSDVFDG---DLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPOP 1137
Db 1282 TGDERSDSSAREVGVLNR-----LDLPVEDDYLMPT-TCQCP 1319
Qy 1138 EYVNPQDVVRPQPSREGPLPAARPAATLERAKTLSPQKNGVVKDVFAFGAVENPEYL 1197
Db 1320 NNNNNNN-----NPNQNNMAAVGAAGYM-----DLIGVPVSDNPEYL 1358
Qy 1198 ----TPQGAAPQPH-----PPAFSP-AFDNLYYWD 1224
Db 1359 LNAQTLGVGESPIQTIGIPVWGEGTWEVKVPMPSSEPTSSDHEYND 1408

```

RESULT 12

```

ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726 (1985).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFLV.
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

```

Query Match 25.7%; Score 1749.5; DB 1; Length 634;
 Best Local Similarity 52.3%; Pred. No. 5.7e-86;
 Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;

```

Qy 587 CAHYKDPPEVCARCPGVKPDLSYMPKFPDEEGACQPCINCTHSCVDLDDKGCPEAQ 646
Db 3 CAHFDIGPHCVKACFAGVLGENDTL-VMKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58
Qy 647 RASPLSTISVAVV-GILLVVVLGVVFGIILKROQKIRKYTMRLRLQETELVEPLTPSGA 705
Db 59 NGSKTPSAAGVVGGLLCLVVGGLGIGLVLRH-HIVKRTLRRLRLQELVEPLTPSGE 117
Qy 706 MPNQAMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANK 765
Db 118 APNQAHRLKETEFKVKVGLSGAFGVYKGLWIPEGEKVKIPVAIKELREATSPKANK 177
Qy 766 EILDEAYVMAGVGSFVYGRLLIGICLTSTVQLVLTQMPYGCCLLDHVRNRRGLSQDLLN 825
Db 178 EILDEAYVMASVDNPHVCRLLIGICLTSTVQLVLTQMPYGCCLLDYIREHNDTNGSYLLN 237
Qy 826 CNQIAKMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVP 885
Db 238 CVQIAKGNVLEERLVRDLAARNVVKTPQHVKITDFGLAKLLGADKEVHAEKGVKVP 297
Qy 886 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLP 945

```

```
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTGSKPYDGIPIASEISSVLEKGERLPQP 357
QY 946 ICTIDVYMTVMKCMWIDSECRPRFRELVSFSESMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYMTVMKCMWIDSECRPRFRELVSFSESMARDPQRFVVIQ-NEGLHPLSPDTSKF 417
QY 1005 YRSLLEDDMDGLVDAAEYLVPOQGFCDPDPAPGAGMWHHRHRSSTSGGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1065 PSEEEAPRSP-----APSEGAGSDVFDGDLGMAAGKQLSLTPHDPSPLOQRYSEPTVP 1119
Db 450 -----SRTPLLSLSLTSATSNNSATNCID-----RNGQGHFVREDTSFQRYSSDPTGN 495
QY 1120 LPSET--DGVVAPLTCSPQEVYQPDVVRPQPPSPREGPLPAARPAGATLERAKTISPGK 1177
Db 496 FLEESIDDOGFL-----PAPEYVNO--LMPKKFS-----TAMVQ 526
QY 1178 NGVVKDOVF-----AFGAVENPEYLTPOGGAAPQPHPPAFSPAFDNLY 1222
Db 527 NQIYNNISUTAIKLPMDSRYSQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSPY 578
QY 1223 WDQ-----DPPE-----RCAPPSTFKGTPTAENPEYLGIDVP 1254
Db 579 WTQSGNHQINLDNPDYQODFLPNETKPNGLLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER STANDARD; PRT; 604 AA.
AC P00535;
CT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erBB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain E54).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erBB gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328659;
RA Debuire B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A06444; TVYUH.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR00719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD00001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MH; 768BCDD06745D609 CRC64;

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 1.6e-83;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

QY 587 CAHYKDPPEVCARCPGKPDLSYMPKFPDEEGACQPCINCHTSCVDLDDKGCQAEQ 646
Db 3 CAHFIDGPHCVKACPAGVIGENDTL-VRYADANAVCOLCHPNCTRGCKPGLEGCP- 58
QY 647 RASPLTISVAVV-GILLVVLGVVFGILIKRRQKIRKYMRRLQETELVEPLTPSGA 705
Db 59 NSKTPSIAAGVVGGLLCLVVGGLGILYLRH-HIVRKRTLRLQLQERLEPLTPSGE 117
QY 706 MPNQAQMRILKETELRKVKVLGSGAGFTYVKGHWIPDGENVKIPVAIKVLENTSPKANK 765
Db 118 APNQAHLRIKETEFKKVKVLGSGAGFTYVKGHWIPDGENVKIPVAIKVLENTSPKANK 177
QY 766 EILDEAYVMAGVGSVYVRLGCLTSTVOLTPMLPYGCLLDHVRNRCRLGSDLLNW 825
Db 178 EILDEAYVMASVDNPHVCRLGLTSTVQLITQLMPYGLLDYIREHKDNIGSQYLLNW 237
QY 826 CMQIAGMSYLEDLVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 885
Db 238 CVQIAGMNYLERRLVRHDLAARNVLVKTPOHKITDFGLAKLLGADEKEYHAEGKVP 297
QY 886 IKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQP 945
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQP 357
QY 946 ICTIDVYMTVMKCMWIDSECRPRFRELVSFSESMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYMTVMKCMWIDSECRPRFRELVSFSESMARDPQRFVVIQ-NEGLHPLSPDTSKF 417
QY 1005 YRSLLEDDMDGLVDAAEYLVPOQGFCDPDPAPGAGMWHHRHRSSTSGGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1065 PSEEEAPRSP-----APSEGAGSDVFDGDLGMAAGKQLSLTPHDPSPLOQRYSEPTVP 1119
Db 450 -----SRTPLLSLSLTSATSNNSATNCID-----RNGQGHFVREDTSFQRYSSDPTGN 495
QY 1120 LPSET--DGVVAPLTCSPQEVYQPDVVRPQPPSPREGPLPAARPAGATLERAKTISPGK 1177
Db 496 FLEESIDDOGFL-----PAPEYVNO--LMPKKPSTAM----- 524
QY 1178 NGVVKDOVF-----AFGAVENPEYLTPOGGAAPQPHPPAFSPAFDNLY 1218
Db 525 --VQYIYNFISLTAISKLPMDSRYSQNSHSTAVDNPEYL-----NTNQSPLAKTVFE 574
QY 1219 NLYYWDQDPPERGAPPSTFKGTPTAENPEY 1248
```

Db	178	ELDEAYVMSVNDPVHVRLLGLICLTSTVQLIQTQLMPYGCGLDIYREHKDNIGSYQLLNW	237
Qy	826	CMQIAKMGSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP	885
Db	238	CVQIAKMGNYLEERHVMVRDLAARNVLKTPQHVKITDFGLAKQLGADKEXYHAEGGKVP	297
Qy	886	IKWMALESILRRPFTHQSDVWSGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPOPP	945
Db	298	IKWMALESILHRIYTHQSDVWSGVTVWELMTGSKPYDGI PASEISSVLEKGERLPOPP	357
Qy	946	ICTIDVYIMVWKMIIDSECRPRFRELVSFEFSMARDPQRFVVIQ-NEDLGPASPLDSTF	1004
Db	358	ICTIDVYIMVWKMSDMSDASRPREFELIAEFSKWARDPPRYLVIQGDERWHLPSTDSKF	417
Qy	1005	YRSLLDDDDGDLVDAAEYLVPOQGFCDPAFCAGGMVHRRHSSTSGGGDLTLGLE	1064
Db	418	YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST-----	449
Qy	1065	PSEEAAPRSL-----APSGAGSDVFDGDLGMAAKQLSLTHDPSPILQRYSEDTVP	1119
Db	450	-----SRTPLLSLSLTSNNSATNCIDRNG-----H-----	476
Qy	1120	LPSETGCVAPLTCSPQPEYVNPQDVVRPQPPSPREGPLPAARPAGAT-LERAKTLPSPGN	1178
Db	477	-PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNIYVILTAISKLPIDRSYQN	527
Qy	1179	GVVKDVFAGGAVENPEYL 1197	
Db	528	-----SHSTAVNDNPEYL 539	
RESULT 15			
EGFR_CHICK			
ID	EGFR_CHICK	STANDARD;	PRT; 703 AA.
AC	P13387;		
DT	01-JAN-1990 (Rel. 13, Last sequence update)		
DT	01-JAN-1990 (Rel. 13, Last sequence update)		
DE	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER		
DE	(Fragment).		
GN	EGFR.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=88261272; PubMed=3260329;		
RA	Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,		
RA	Ullrich A., Vennstrom B., Schlessinger J., Givol D.,		
RT	"Chicken epidermal growth factor (EGF) receptor: cDNA cloning,		
RT	expression in mouse cells, and differential binding of EGF and		
RT	transforming growth factor alpha.";		
RL	Mol. Cell. Biol. 8:1970-1978(1988).		
CC	!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,		
CC	AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND		
CC	VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).		
CC	!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein		
CC	tyrosine phosphate.		
CC	!- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	!- MISCELLANEOUS: Binding of EGF to the receptor leads to		
CC	dimerization, internalization of the EGF-receptor complex,		
CC	induction of the tyrosine kinase activity, stimulation of cell DNA		
CC	synthesis, and cell proliferation.		
CC	!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/).		

or send an email to license@sib.ch).

```
CC EMBL; M20386; AAA48760.1;
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; PARTIAL.
DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 667 POTENTIAL.
FT DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL).
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 201 214 BY SIMILARITY.
FT DISULFID 222 230 BY SIMILARITY.
FT DISULFID 226 238 BY SIMILARITY.
FT DISULFID 239 247 BY SIMILARITY.
FT DISULFID 243 255 BY SIMILARITY.
FT DISULFID 258 267 BY SIMILARITY.
FT DISULFID 271 298 BY SIMILARITY.
FT DISULFID 302 314 BY SIMILARITY.
FT DISULFID 318 333 BY SIMILARITY.
FT DISULFID 336 340 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 581 BY SIMILARITY.
FT DISULFID 569 589 BY SIMILARITY.
FT DISULFID 592 601 BY SIMILARITY.
FT DISULFID 605 627 BY SIMILARITY.
FT DISULFID 630 638 BY SIMILARITY.
FT DISULFID 634 646 BY SIMILARITY.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 703 703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;
```

Query Match 23.4%; Score 1595; DB 1; Length 703;

Best Local Similarity 44.6%; Pred. No. 1e-77;

Matches 316; Conservative 111; Mismatches 251; Indels 30; Gaps 14;

```
QY 8 RAGLLALLPPGAA-----STQVCTGTDMLKLPASPETHLDMLRHLYQGGCVQGNLE 61
DB 13 RGAALVLLLLGVALCSAVEEKKVCGGTNNKLTQLGHVEDFTSLQRMYNCCVWLSNLE 72
QY 62 LTYLPTNASLFDIQEYQGVYLIANQVQVQLQRLIRIVRGTLQFDNYALAVLDNGD 121
DB 73 ITYVEHNRLDTLTKTIQEVAGYVLIANVVDVIPLENLQIIRGNVLYDNSFALAVLSNVH 132
QY 122 PLNNTTPVTGASPGGLRELQLSLTEILKGVLIQRNPOLCYODTILWKDIIFHKNQLAL 181
DB 133 -NNKTO-----GLRELPMKRLSELNGVVKISNNPKLCNNDTVLWNDIIDSRSK-PL 182
QY 182 TLID-TNRSRACHPCSPCKSGRCWSESSEDQSLTRTVAGGCA-RCKGPLPTDCCHEQ 239
```

```
DB 183 TVLDFASNLSSCPKCHPNCTEDHMGAGEBQNCOTLTKVICAQCSGRCGRKVPSPDCCHNQ 242
QY 240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTPESPNPEGRVTFGASCVTAC 299
DB 243 CAAGCTGPRSDCLACRKRFRDATTCTPPLVLYNPTTYQMDVNPGEKTSFGATCVREC 302
QY 300 PYNILSTDVGSCTLVCPLNHNEVTAEDGTQRCCEKSCPCARVCYGLGMOYIKANSKFIGI 359
DB 303 PHNYVVDHSGSVRSCTNTDTEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGILS-INA 360
QY 360 TELE-PAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLBEITGYLYISAMPDS 418
DB 361 TNIDSFKNCTKINGDVSIPLVAFGLDAFTKTLPLDPKLDVFRFVKEISGFLLIQAMPDN 420
QY 419 LPDLSVFQNLQVIRGRIHLHGAYSLLTLOGLSWLGSLRELGLALIHNNHLCFVH 478
DB 421 ATDLVAFENLEIRGRTKHQYSLAVNVLKIQSLRLSLKEISDGDIAIMKNKLCYAD 480
QY 479 TYPWDOLFERNPHQALLHTANRPEDECVGEGELACHQLCARGHCWPGPTQCVNCSQFLRGQ 538
DB 481 TNMRSLFATQSKTKIIQNRNKNDCDADRHVCDPLCSDVGCWGPFPFHCFSCRFPSROK 540
QY 539 ECVEECRVLOGLPREVYNARHCLPCHPECOQNG-----SVTCFGEADQCVACAHYKDPFP 595
DB 541 ECVKQCNILQGEPRFEPERDSKCLPCHSECLVNSTAYNTTCSGPGPDHCKCAHFDGPH 600
QY 596 CVARCPGKVPDLSPYMPIWKFPDEGACOPCPINCTHSCVDLDDKGPABQORASPLTSIV 655
DB 601 CVKACPAVGLGENDTL-VMKYADANAVCOLCHPNCTRGCCKGPGLEGCP---NGSKTPSIA 656
QY 656 SAVV-GILLVVVGLVVVFGILIKRRQOKIRKTYMRRLLOETELVEPLTP 702
DB 657 AGVVGSLCLVVVGLGILGLYLRRL-HIVKRTLRLLQRELVEPLTP 703
```

Search completed: July 22, 2003, 08:45:48

Job time : 20.2304 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-103-117-14
Perfect score: 6847
Sequence: 1 MELAALCRWGLLLALLPPCA.....TFKGTPTAENPEYGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

```

1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6714	98.1	1255	21	Human heregulin 2
2	6714	98.1	1255	22	Human heregulin 2
3	6714	98.1	1255	22	Human heregulin 2
4	6714	98.1	1255	22	Human heregulin 2
5	6708	98.0	1255	17	Human heregulin 2
6	6708	98.0	1255	20	Human heregulin 2
7	6708	98.0	1255	21	Human heregulin 2
8	6708	98.0	1255	21	Human heregulin 2
9	6708	98.0	1255	22	Human heregulin 2
10	6708	98.0	1255	22	Human heregulin 2

11	6708	98.0	1255	23	AAE24067	Human Her-2 protei
12	6708	98.0	1255	23	AAE24079	Human Her-2/neu pr
13	6708	98.0	1255	23	AAE24079	Human Her-2/neu pr
14	6708	98.0	1255	23	AAU77114	Human Her-2/neu po
15	6665	97.3	1433	14	AAU77114	Sequence of c-erbB
16	6544	95.6	1433	14	AAU77114	Human breast cance
17	6391	93.3	1200	21	AAU98923	Human HER-2/neu pr
18	5950.5	86.9	1256	21	AAU98923	Rat HER-2/neu pr
19	5950.5	86.9	1256	21	AAU98923	Rat HER-2/neu pr
20	5918.5	86.4	1256	21	AAU98923	Mouse Her-2/neu pr
21	5918.5	86.4	1256	21	AAU98923	Mouse Her-2/neu pr
22	5918.5	86.4	1256	21	AAU98923	Amino acid sequenc
23	4794	70.0	919	21	AAU98923	Human HER-2/neu on
24	4794	70.0	919	21	AAU98923	Human HER-2/neu fu
25	4068.5	59.4	920	23	AAU98923	Human HER-2/neu ex
26	4068.5	59.4	920	23	AAU98923	Mouse Her-2/neu ex
27	3678	53.7	712	21	AAU98923	Human HER-2/neu fu
28	3678	53.7	712	21	AAU98923	Human HER-2/neu fu
29	3532	51.6	782	18	AAU98923	Her-2/neu extracel
30	3530	51.6	782	18	AAU98923	Her-2/neu extracel
31	3530	51.6	782	18	AAU98923	Extracellular HER-
32	3492	51.0	645	22	AAU98923	Human ErbB2 oncopr
33	3492	51.0	645	22	AAU98923	Human ErbB2 oncopr
34	3427	50.1	951	21	AAU98923	DC9cFv-erbB2EC fu
35	3326	48.6	624	11	AAU98923	Extracellular port
36	3139	45.8	1210	21	AAU98923	Amino acid sequenc
37	3139	45.8	1210	21	AAU98923	Human EGF receptor
38	3139	45.8	1210	21	AAU98923	Human EGF receptor
39	3139	45.8	1210	21	AAU98923	Human Her-1 protei
40	3137	45.8	1210	22	AAU98923	Human epidermal gr
41	3098	45.2	1210	22	AAU98923	Amino acid sequenc
42	3084	45.0	583	23	AAU98923	Human epidermal gr
43	3084	45.0	583	23	AAU98923	Human protein for
44	3083	45.0	589	23	AAU98923	Human protein for
45	3083	45.0	600	23	AAU98923	Human protein for

ALIGNMENTS

RESULT 1
AAU92620
ID AAY92620 standard; Protein; 1255 AA.
XX AC AAY92620;
XX 10-AUG-2000 (first entry)
XX Human heregulin 2 (Her2).
DE Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX Homo sapiens.
OS Homo sapiens.
XX Key
XX Domain
FT Location/Qualifiers
FT /label= N-terminal
FT /note= "mature polypeptide"
FT 5..25
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 59..73
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 103..117
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 149..163
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 174..323

FT Region /label= Cysteine_rich_domain
 210..224
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 FT Region 230..264
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 FT Domain 324..483
 /label= Ligand_binding_domain
 325..339
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 FT Region 369..383
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 FT Region 465..479
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 FT Domain 484..623
 /label= Cysteine_rich_domain
 579..593
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 FT Domain 624..654
 /label= Transmembrane_domain
 632..652
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 FT Region 653..667
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 FT Domain 655..1010
 /label= Tyrosine_kinase_domain
 661..675
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 FT Region 695..709
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 FT Region 710..730
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 FT Domain 1011..1235
 /label= C-terminal_domain
 W0200020027-A2.
 13-APR-2000.
 05-OCT-1999; 99WO-DK00525.
 05-OCT-1998; 98DK-0001261.
 20-OCT-1998; 98US-0105011.
 (NEBI-) M & E BIOTECH AS.
 Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 Gautam A, Birk P, Karlsson G;
 WPI: 2000-349917/30.
 N-PSDB; AAA09455.
 Inducing immune responses to weakly immunogenic, tumor associated
 peptide antigens for the treatment of breast and prostate cancer
 Claim 62; Page 193-198; 220pp; English.
 This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 Her2 can be used in the claimed method as an autovaccine to induce a CTL
 response. Subdominant CTL epitopes, antibody binding regions and
 cysteine residues involved in disulfide bonds are preserved in the
 immunogenized forms. Regions suitable for the insertion of foreign T
 helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 XX
 SQ Sequence 1255 AA;
 Query Match 98.1%; Score 6714; DB 21; Length 1255;
 Best Local Similarity 98.3%; Pred. No. 0;
 Matches 1243; Conservative 0; Mismatches 8; Indels 14; Gaps 2;
 QY 1 MELAALCRWGLLALLPPGAASTOCTGDMKRLPASPEHLDMLRHLYQGCQVQGNL 60
 DB 1 MELAALCRWGLLALLPPGAASTOCTGDMKRLPASPEHLDMLRHLYQGCQVQGNL 60
 QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHNOVQVPLQRLRIVRGTOIFEDNYALAVLDNG 120
 DB 61 ELTYLPTNASLSFLQDIQEVQGVLIHNOVQVPLQRLRIVRGTOIFEDNYALAVLDNG 120
 QY 121 DPLNNFNFTVSFMLRVPKVSAS----HLEQLRSLTEILKGGVLIQRNPOLCYQDTILWK 176
 DB 121 DPLNN-----TTPVTGASPGGLRQLRSLTEILKGGVLIQRNPOLCYQDTILWK 170
 QY 177 DIFHKNNQALALTIDNRSRACHPCSPCKGSRCESSDCCSLTRTVCAGCARCKGP 236
 DB 171 DIFHKNNQALALTIDNRSRACHPCSPCKGSRCESSDCCSLTRTVCAGCARCKGP 230
 QY 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYT 296
 DB 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYT 290
 QY 297 FGASCVTACPYNYLSTDVGSCTLVCPHNOEVTAEQGTORCEKCKSPCARVCYGLGMEHL 356
 DB 291 FGASCVTACPYNYLSTDVGSCTLVCPHNOEVTAEQGTORCEKCKSPCARVCYGLGMEHL 350
 QY 357 REVRAVTSANIOEFAGCKXIFGSLAPLSPESFGDPPASNTAPLOPEQLQVFEETITGYL 416
 DB 351 REVRAVTSANIOEFAGCKXIFGSLAPLSPESFGDPPASNTAPLOPEQLQVFEETITGYL 410
 QY 417 YISAWPDSLPLDSVFONLQVIRGRILHNGAYSILTLOGLGISWGLRSLRSLGSLALIH 476
 DB 411 YISAWPDSLPLDSVFONLQVIRGRILHNGAYSILTLOGLGISWGLRSLRSLGSLALIH 470
 QY 477 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEBCEVGEGLACHOLCARGHCWGPGTQCVN 536
 DB 471 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEBCEVGEGLACHOLCARGHCWGPGTQCVN 530
 QY 537 CSQFLRGQECVBECEVQLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHY 596
 DB 531 CSQFLRGQECVBECEVQLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHY 590
 QY 597 KDPFFCVARCPGKVPDLSYMPIWKFPDDEGACQPCPINCTHSCVDLDDKGCFAEORASP 656
 DB 591 KDPFFCVARCPGKVPDLSYMPIWKFPDDEGACQPCPINCTHSCVDLDDKGCFAEORASP 650
 QY 657 LTSIVSAVVGILLVVVLGVVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNOA 716
 DB 651 LTSIVSAVVGILLVVVLGVVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNOA 710
 QY 717 QMRILKETELRKVKVLGSAFGTVYKGIWIPGENYKIPVAIKVLRNTPSPANKBILDE 776
 DB 711 QMRILKETELRKVKVLGSAFGTVYKGIWIPGENYKIPVAIKVLRNTPSPANKBILDE 770

QY 777 AYVMAGVSPYVSRLLGICLTSTVOLVTPQMPYGCLLDHVRENRRGLSQDILLNWCMTA 836
 DB 771 AYVMAGVSPYVSRLLGICLTSTVOLVTPQMPYGCLLDHVRENRRGLSQDILLNWCMTA 830
 QY 837 KGMYSLEVDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWA 896
 DB 831 KGMYSLEVDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWA 890
 QY 897 LESILRRRFTHOSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPICTID 956
 DB 891 LESILRRRFTHOSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPICTID 950
 QY 957 VYIMVKKWIMIDSECRPRFRELVSFMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLE 1016
 DB 951 VYIMVKKWIMIDSECRPRFRELVSFMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLE 1010
 QY 1017 DDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEA 1076
 DB 1011 DDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEA 1070
 QY 1077 PRSPLAPSEGAGSDVFDGLGMAAKGLOSLTPHPSPLQRYSEPTVPLPSETDGYVAP 1136
 DB 1071 PRSPLAPSEGAGSDVFDGLGMAAKGLOSLTPHPSPLQRYSEPTVPLPSETDGYVAP 1130
 QY 1137 LTCSPQPEVNOPVRRPQPPSPREGPLPAARAGATLERAKTSLSPGKNGVKDVFAGGA 1196
 DB 1131 LTCSPQPEVNOPVRRPQPPSPREGPLPAARAGATLERAKTSLSPGKNGVKDVFAGGA 1190
 QY 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
 DB 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250
 QY 1257 LDVPV 1261
 DB 1251 LDVPV 1255

RESULT 2
 AAEL12130
 ID AAEL12130 standard; Protein; 1255 AA.
 AC AAEL12130;
 DT 18-DEC-2001 (first entry)
 DE Human tyrosine kinase-type receptor, HER-2.
 KW Therapeutic compound; major histocompatibility complex; vaccine;
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
 KW antigen presenting cell; human; tyrosine kinase-type receptor.
 OS Homo sapiens.
 FT Key Location/Qualifiers
 FT Region 774..782
 FT /note= "Antigenic epitope"
 XX
 FN WO200168677-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 16-MAR-2001; 2001WO-US40328.
 XX
 PR 16-MAR-2000; 2000US-0527487.
 XX
 PA (GENZ) GENZYME CORP.
 XX
 PI Nicolette CA;
 XX
 DR WPI; 2001-616284/71.
 DR N-PSDB; AAD19731.
 XX

PT Novel synthetic therapeutic compound for inducing immune response and
 PT for use in adoptive immunotherapy, has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties
 XX
 PS Claim 4; Page 63-67; 69pp; English.
 XX
 CC The invention relates to synthetic therapeutic compounds (antigenic
 CC peptides) with enhanced binding to major histocompatibility complex
 CC (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterised by expression of the breast cancer antigen, HER-2.
 CC Polynucleotides that encode peptides of the invention are useful as
 CC hybridisation probes and as primers for the detection of genes of gene
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 CC (774-782).
 XX
 SQ Sequence 1255 AA;
 Query Match 98.1%; Score 6714; DB 22; Length 1255;
 Best Local Similarity 98.3%; Pred. No. 0;
 Matches 1243; Conservative 0; Mismatches 8; Indels 14; Gaps 2;
 QY 1 MELAALCERWGLLLALLPGAASTQVCTGTDKMLRLPASPETHLDMRLHYQSCVVQGNL 60
 DB 1 MELAALCERWGLLLALLPGAASTQVCTGTDKMLRLPASPETHLDMRLHYQSCVVQGNL 60
 QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIANQVQVPLQRLRIVRGTQLPEDNYALAVLDNG 120
 DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIANQVQVPLQRLRIVRGTQLPEDNYALAVLDNG 120
 QY 121 DPLNNFNFTVFWLRVPKVSAS-----HLEQLRSLTEILKGGVLIQRNPQLCYQDTILWK 176
 DB 121 DPLNN-----TTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWK 170
 QY 177 DIFHKNQALTLTDNRSRACHPCSPKSGRCSGESSEDCOSLTRTVCAGGCARCKGP 236
 DB 171 DIFHKNQALTLTDNRSRACHPCSPKSGRCSGESSEDCOSLTRTVCAGGCARCKGP 230
 QY 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRYT 296
 DB 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRYT 290
 QY 297 FGASCVTACPNYLSLTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLMEHL 356
 DB 291 FGASCVTACPNYLSLTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLMEHL 350
 QY 357 REVRAVTSANTIOEFAGCKKIFGSLAFLPESDGDPSANTAPLQEPQLVFTEETGYL 416
 DB 351 REVRAVTSANTIOEFAGCKKIFGSLAFLPESDGDPSANTAPLQEPQLVFTEETGYL 410
 QY 417 YISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGURSRELGSGLALIH 476
 DB 411 YISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGURSRELGSGLALIH 470
 QY 477 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHMGPGTQCVN 536
 DB 471 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHMGPGTQCVN 530
 QY 537 CSQFLRGQECVEECRVLOGLPREYNAHCLPCHPECPONGSVTCFGEADQCACAHY 596
 DB 531 CSQFLRGQECVEECRVLOGLPREYNAHCLPCHPECPONGSVTCFGEADQCACAHY 590
 QY 597 KDPFFCVARCPGKVPKPLSYMPIWKFPDEEGACOPCPINCTHSCVDLDDKGPAPORASP 656
 DB 591 KDPFFCVARCPGKVPKPLSYMPIWKFPDEEGACOPCPINCTHSCVDLDDKGPAPORASP 650

QY 657 LTSISAVVGIILVVVLGVVFGILIKRROOKIRKVTMRRLLOETELVEPLTPSGAMPNOA 716
DB 651 LTSISAVVGIILVVVLGVVFGILIKRROOKIRKVTMRRLLOETELVEPLTPSGAMPNOA 710
QY 717 QMRILKETELRKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDE 776
DB 711 QMRILKETELRKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDE 770
QY 777 AVVMAGVSPYVSRLIGICTSTVOLVTOLMPYGCLLDHHVNRGRGLGSODLLNMCQIA 836
DB 771 AVVMAGVSPYVSRLIGICTSTVOLVTOLMPYGCLLDHHVNRGRGLGSODLLNMCQIA 830
QY 837 KCMYSLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVPWKMA 896
DB 831 KCMYSLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVPWKMA 890
QY 897 LESILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 956
DB 891 LESILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 950
QY 957 VYMIWVKMIDSECRPRELVSFSEMRARDPQRFVJQNEDLGPASPLDSTFYRSLLE 1016
DB 951 VYMIWVKMIDSECRPRELVSFSEMRARDPQRFVJQNEDLGPASPLDSTFYRSLLE 1010
QY 1017 DDDMGDLVDAEYLVPOQGFCCPDPAAGAGGMVHHRSSSTRSGGGDLTLGLEPSEEEA 1076
DB 1011 DDDMGDLVDAEYLVPOQGFCCPDPAAGAGGMVHHRSSSTRSGGGDLTLGLEPSEEEA 1070
QY 1077 PRSPLAPSEGAGSDVFDGLGNAAGLQSLPTHDPSPLQRYSEDPVPLPSETDGYVAP 1136
DB 1071 PRSPLAPSEGAGSDVFDGLGNAAGLQSLPTHDPSPLQRYSEDPVPLPSETDGYVAP 1130
QY 1137 LTCSPQPEYVNPQDVRPOPSPREGPLPAARAGATLERAKTSLSPCKNGVVKDVFAGGA 1196
DB 1131 LTCSPQPEYVNPQDVRPOPSPREGPLPAARAGATLERAKTSLSPCKNGVVKDVFAGGA 1190
QY 1197 VENPEYLTPOGGAAPOPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
DB 1191 VENPEYLTPOGGAAPOPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250
QY 1257 LDVPV 1261
DB 1251 LDVPV 1255

RESULT 3
AAB60167

ID AAB60167 standard; Protein; 1255 AA.

AC AAB60167;

DT 03-APR-2001 (first entry)

DE HER2 transgene plasmid construct encoded protein.

KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
antibody.

XX Homo sapiens.

OS Synthetic.

XX WO200100244-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-US17229.

XX 25-JUN-1999; 99US-0141316.

PR 16-MAR-2000; 2000US-0189844.

XX (GTEII) GENENTECH INC.

XX Erickson S, Schwall R;

PI

XX
DR
DR
XX
PT
PT
PT
XX
PS
XX
CC
CC
CC
CC
CC
CC
XX
SQ

WPI: 2001-061962/07.
N-PSDB; AAF24297.
Treating tumors, particularly breast cancers, which overexpress an ErbB receptor and does not respond to an anti-ErbB antibody, comprises conjugating the antibody to a maytansinoid -
Example 3; Fig 4; 92pp; English.
The present invention provides a method of treating cancer by administering a conjugate of anti-ErbB antibody with a maytansinoid. In particular, the antibody is directed against ErbB2 (also known as HER2 and p185neu). The method is particularly useful in the treatment of breast, ovarian, stomach, endometrial, salivary gland, lung, kidney, colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
Query Match 98.1%; Score 6714; DB 22; Length 1255;
Best Local Similarity 98.3%; Pred No. 0;
Matches 1243; Conservative 0; Mismatches 8; Indels 14; Gaps 2;
QY 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPEHLDMLRLHYQGCQVVGNL 60
DB 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPEHLDMLRLHYQGCQVVGNL 60
QY 61 EUTYLPNASLSFLODIOEVQYVLIHNVQVPLQRLRIRVGTQLFEDNYALAVLDNG 120
DB 61 EUTYLPNASLSFLODIOEVQYVLIHNVQVPLQRLRIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNFNFTVSFMLRVPKVSAS----HLEQLRSLEILKGGVLIORNPOLCYODTILWK 176
DB 121 DPLNN-----TTPVTGASPGGLRELQRLSLEILKGGVLIORNPOLCYODTILWK 170
QY 177 DIFHKNNQALALTLIDNRSRACHPCSPCKSGRCWGESSEDCOSLTRTVCAGGCARCKGP 236
DB 171 DIFHKNNQALALTLIDNRSRACHPCSPCKSGRCWGESSEDCOSLTRTVCAGGCARCKGP 230
QY 237 LPTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 296
DB 231 LPTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 290
-/- 297 FGASCVTACPNYLSLTDVGSCTLVCPLNHNVTAEDGTORCEKCKPCARVCYGLGMEHL 356
DB 291 FGASCVTACPNYLSLTDVGSCTLVCPLNHNVTAEDGTORCEKCKPCARVCYGLGMEHL 350
QY 357 REVRAVTSANIOEFAGCKKIFGSLAPLPSFDGPASNTAPLOPEQLQVFETLEEITGYL 416
DB 351 REVRAVTSANIOEFAGCKKIFGSLAPLPSFDGPASNTAPLOPEQLQVFETLEEITGYL 410
QY 417 YISAWPDSLPLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLGSLALIH 476
DB 411 YISAWPDSLPLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLGSLALIH 470
QY 477 NTHLCFVHTVPWDQLFRNPHQALLHTANRPECVCGEGLACHOLCARGHCWGPGTQCVN 536
DB 471 NTHLCFVHTVPWDQLFRNPHQALLHTANRPECVCGEGLACHOLCARGHCWGPGTQCVN 530
QY 537 CSQFLRGQCEVBECEVRLQGLPREYVNAHCLPCHPECQPNQSVTCFGEADOCVACAHY 596
DB 531 CSQFLRGQCEVBECEVRLQGLPREYVNAHCLPCHPECQPNQSVTCFGEADOCVACAHY 590
QY 597 KDPFPCVAPCGVGPDLISYMPIWKPDPBEGACQPCPINCTHSCVLDLDDKGCPEORASP 656
DB 591 KDPFPCVAPCGVGPDLISYMPIWKPDPBEGACQPCPINCTHSCVLDLDDKGCPEORASP 650
QY 657 LTSISAVVGIILVVVLGVVFGILIKRROOKIRKVTMRRLLOETELVEPLTPSGAMPNOA 716
DB 651 LTSISAVVGIILVVVLGVVFGILIKRROOKIRKVTMRRLLOETELVEPLTPSGAMPNOA 710
QY 717 QMRILKETELRKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDE 776

Db 711 QNRILKETELRKVKVVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770
QY 777 AYVMAGVGSPPYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRRGLSGODLLNWCQIA 836
Db 771 AYVMAGVGSPPYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRRGLSGODLLNWCQIA 830
QY 837 KGSYLEDLVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHYHADGGKVPKQMA 896
Db 831 KGSYLEDLVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHYHADGGKVPKQMA 890
QY 897 LESILRRRTHOSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 956
Db 891 LESILRRRTHOSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 950
QY 957 VYMIWVKWMIIDSECRPRRELVSERPMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLLE 1016
Db 951 VYMIWVKWMIIDSECRPRRELVSERPMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLLE 1010
QY 1017 DDDMGDLVDAEYLVLPQOQFFCPDPAPGAGGMVHRRSSSTRSGGGDLTLGLEPSEEA 1076
Db 1011 DDDMGDLVDAEYLVLPQOQFFCPDPAPGAGGMVHRRSSSTRSGGGDLTLGLEPSEEA 1070
QY 1077 PRSPLAPSGAGSDVDGDLGMAAGLQSLPHTDPSPLQRYSEDTVPPLPSETDGYVAP 1136
Db 1071 PRSPLAPSGAGSDVDGDLGMAAGLQSLPHTDPSPLQRYSEDTVPPLPSETDGYVAP 1130
QY 1137 LTCSPQEVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVWVDVAFGGA 1196
Db 1131 LTCSPQEVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVWVDVAFGGA 1190
QY 1197 VENPEYLTQGGAAPOHPHPAPSPAFDNLVYWDQPPRGAPPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTQGGAAPOHPHPAPSPAFDNLVYWDQPPRGAPPSTFKGTPTAENPEYLG 1250
QY 1257 LDVVPV 1261
Db 1251 LDVVPV 1255

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX AAU74545;
AC AAU74545;
XX
XX 23-APR-2002 (first entry)
XX
DE Human HER2 (ErbB2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoealic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
XX US2002001587-A1.
XX
XX 03-JAN-2002.
XX
XX 16-MAR-2001; 2001US-0811123.
XX
XX 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
XX (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWKOWSKI M.
XX

PI Erickson S, Schwall R, Sliwowski M;
XX WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
PT Treating tumour characterised by overexpression of epidermal growth
factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
XX
PS Example 3; Fig 7; 93pp; English.
XX
CC The invention relates to treating a tumour in a mammal, where the tumour
is characterised by the overexpression of an epidermal growth factor
receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;
Query Match 98.1%; Score 6714; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1243; Conservative 0; Mismatches 8; Indels 14; Gaps 2;
QY 1 MELAALCRNGLLLALLPAGAASQVCTGTDMLRLPASPEHLDMRLHLYQGCVQVGNL 60
Db 1 MELAALCRNGLLLALLPAGAASQVCTGTDMLRLPASPEHLDMRLHLYQGCVQVGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHQNVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVLIHQNVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120
QY 121 DPLNPNFTVFWLVRPKVSAS-----HLEQLRSLTEILKGGVLIQRNPOLCYQDTTLWK 176
Db 121 DPLNN-----TTPVTGASPGGLRELRLSLTEILKGGVLIQRNPOLCYQDTTLWK 170
QY 177 DIFHKNNQALALTIDTNRSRACHPCSPMCKGRGWESSEDCOSLRTVTCAGGCARCKGP 236
Db 171 DIFHKNNQALALTIDTNRSRACHPCSPMCKGRGWESSEDCOSLRTVTCAGGCARCKGP 230
QY 237 LPTDCCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 296
Db 231 LPTDCCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290
QY 297 FGASCVTACPNYVLTVDVGSCTLVCPHMQEVTAEQGTORCEKCKSPCARVCYGLGMEHL 356
Db 291 FGASCVTACPNYVLTVDVGSCTLVCPHMQEVTAEQGTORCEKCKSPCARVCYGLGMEHL 350
QY 357 REVRAVTSANIOBFAGCKKIFGSLAPLPESFGDPSANTAPLOPEOLOVETLEEITGYL 416
Db 351 REVRAVTSANIOBFAGCKKIFGSLAPLPESFGDPSANTAPLOPEOLOVETLEEITGYL 410
QY 417 YISAWPDSLPLDLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSLRGLSGLALIH 476
Db 411 YISAWPDSLPLDLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSGLSGLALIH 470
QY 477 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEBECVGEGLACHOLCARGHCWGPGTQCVN 536
Db 471 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEBECVGEGLACHOLCARGHCWGPGTQCVN 530
QY 537 CSQFLRGQECVEECVRLQGLPREYNARHCLPCHPECPQNGSVTCFGEADOCVACAHY 596
Db 531 CSQFLRGQECVEECVRLQGLPREYNARHCLPCHPECPQNGSVTCFGEADOCVACAHY 590
QY 597 KDPPFCVARCPGSKVVKPDLISYMPDWKPPDEEGACQPCPINCTHSCVDLDDKGCPEQRASP 656


```
QY 597 KDPPECVARGSGVXPDLSPYPIWKFPDDEGACQPCPINCTHSCVDLDDKGCFAEORASP 656
Db |||||
QY 591 KDPPECVARGSGVXPDLSPYPIWKFPDDEGACQPCPINCTHSCVDLDDKGCFAEORASP 650
Db |||||
QY 657 LTSIVSAVGIILLVVVLGVVFGILIKRROQKIRKKTMRLLQETELVEPLTPSGAMPNOA 716
Db |||||
QY 651 LTSIVSAVGIILLVVVLGVVFGILIKRROQKIRKKTMRLLQETELVEPLTPSGAMPNOA 710
Db |||||
QY 717 QMRILKETELRKVKVLSGAGFTVYKGIWIPGENVKIPVAIKVLRNTPSKANKEILDE 776
Db |||||
QY 711 QMRILKETELRKVKVLSGAGFTVYKGIWIPGENVKIPVAIKVLRNTPSKANKEILDE 770
Db |||||
QY 777 AYVMAGVSPYVSRLLIGLCTSTVOLVTOLMPYGCLLDHVNRGRGLSGQDLLNWCMTA 836
Db |||||
QY 771 AYVMAGVSPYVSRLLIGLCTSTVOLVTOLMPYGCLLDHVNRGRGLSGQDLLNWCMTA 830
Db |||||
QY 837 KGMYSLEVDRLVHROLAARNVLVKSPPNHVKITDFGLARLLDIDETEHADGKVPWKMA 896
Db |||||
QY 831 KGMYSLEVDRLVHROLAARNVLVKSPPNHVKITDFGLARLLDIDETEHADGKVPWKMA 890
Db |||||
QY 897 LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTID 956
Db |||||
QY 891 LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTID 950
Db |||||
QY 957 VYIMVWKMWIDSECRPRELVSFSESRMARDPQRFVWQNEIDLGPASPLDSTFYRSLE 1016
Db |||||
QY 951 VYIMVWKMWIDSECRPRELVSFSESRMARDPQRFVWQNEIDLGPASPLDSTFYRSLE 1010
Db |||||
QY 1017 DDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEA 1076
Db |||||
QY 1011 DDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEA 1070
Db |||||
QY 1077 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLTPHDPSPLOQRYSEDPVPLPSETDGYVAP 1136
Db |||||
QY 1071 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLTPHDPSPLOQRYSEDPVPLPSETDGYVAP 1130
Db |||||
QY 1137 LTCSPQPEVNPQDVRPOPSPREGPLPAARAGATLERAKTLLSPGKNGVVKDVFAGGA 1196
Db |||||
QY 1131 LTCSPQPEVNPQDVRPOPSPREGPLPAARAGATLERAKTLLSPGKNGVVKDVFAGGA 1190
Db |||||
QY 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLG 1256
Db |||||
QY 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLG 1250
Db |||||
QY 1257 LDVPV 1261
Db |||||
QY 1251 LDVPV 1255
Db |||||

RESULT 6
AAW92406
ID AAW92406 standard; Protein; 1255 AA.
XX
AC AAW92406;
XX
DT 21-APR-1999 (first entry)
XX
DE Human HER-2/neu oncogene protein.
XX
KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
XX malignancy; treatment; tumour.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 676..1255
FT /note= "region which elicits immune response"
XX
PN US5869445-A.
XX
PD 09-FEB-1999.
XX
PF 01-APR-1996; 96US-0625101.
```

```
XX
PR 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
PA (UNIW ) UNIV WASHINGTON.
XX
PI Cheever MA, Disis ML;
XX
DR WPI: 1999-152835/13.
DR N-PSDB; AAX01912.
XX
PT Use of HER-2/neu polypeptides - for eliciting an immune response to
PT an HER-2/neu associated malignancy, particularly for treating or
PT preventing tumours
XX
PS Claim 3; Column 31-38; 26pp; English.
XX
CC This sequence represents the human HER-2/neu oncogene protein. A fragment
CC of this protein is used in a method for eliciting or enhancing an immune
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.
XX
SQ Sequence 1255 AA;

Query Match 98.0%; Score 6708; DB 20; Length 1255;
Best Local Similarity 98.1%; Pred. No. 0;
Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;

QY 1 MELAALCRWGLLLALLPGGAASQVCTGDMKRLPASPEHLDMRLHLYQGCQVVOGNL 60
Db |||||
QY 1 MELAALCRWGLLLALLPGGAASQVCTGDMKRLPASPEHLDMRLHLYQGCQVVOGNL 60
Db |||||
QY 61 ELTYLPTNASLSFLQDIOEVQGVLIHNVQVPLQRLRIRVGTQLFEDNALAVLDNG 120
Db |||||
QY 61 ELTYLPTNASLSFLQDIOEVQGVLIHNVQVPLQRLRIRVGTQLFEDNALAVLDNG 120
Db |||||
QY 121 DPLNPNFTVSFWLRPKVSAS----HLEQLRSUTEILKGGVLIQRNPOLCYQDTILWK 176
Db |||||
QY 121 DPLNPNFTVSFWLRPKVSAS----HLEQLRSUTEILKGGVLIQRNPOLCYQDTILWK 170
Db |||||
QY 177 DIFHKNQALTLIDTNESRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGGCARCKGP 236
Db |||||
QY 171 DIFHKNQALTLIDTNESRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGGCARCKGP 230
Db |||||
QY 237 LPTDCCHQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 296
Db |||||
QY 231 LPTDCCHQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290
Db |||||
QY 297 FGASCVTACPNYLSLTDVGSCTLVCPLNQNEVTAEDGTORCEKCKPCARVCYGLQMEHL 356
Db |||||
QY 291 FGASCVTACPNYLSLTDVGSCTLVCPLNQNEVTAEDGTORCEKCKPCARVCYGLQMEHL 350
Db |||||
QY 357 REVRAVTSANTQEPAGCKKIFGSLAPLPESDPDASNTAPLOPQLOVFETLEETGYL 416
Db |||||
QY 351 REVRAVTSANTQEPAGCKKIFGSLAPLPESDPDASNTAPLOPQLOVFETLEETGYL 410
Db |||||
QY 417 YISAWPDSLPDLVSFQNLQVIRGRILHNGAYSLTLQGLGISMLGLRSRELGSGLALIIH 476
Db |||||
QY 411 YISAWPDSLPDLVSFQNLQVIRGRILHNGAYSLTLQGLGISMLGLRSRELGSGLALIIH 470
Db |||||
QY 477 NTHLCFVHTVPWDQLFRPNHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVN 536
Db |||||
QY 471 NTHLCFVHTVPWDQLFRPNHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVN 530
Db |||||
QY 537 CSQFLRGQECVEECRVLQGLPREYNARHCLPCHPECOPOGNSVTCFGEADQCACAHY 596
Db |||||
QY 531 CSQFLRGQECVEECRVLQGLPREYNARHCLPCHPECOPOGNSVTCFGEADQCACAHY 590
Db |||||
```

QY	597	KDPPFCVACPSGVKPDLSYMPYIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRAS	656
Db	591	KDPPFCVACPSGVKPDLSYMPYIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRAS	650
QY	657	LTSIVSAVGIILVVVLGVVFGILIKRQOKIRKYTMRRLLQSTELVBLTPSGAMPNOA	716
Db	651	LTSIIISAVGIIILVVVLGVVFGILIKRQOKIRKYTMRRLLQSTELVBLTPSGAMPNOA	710
QY	717	QNRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKIILDE	776
Db	711	QNRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKIILDE	770
QY	777	AYVMAGVSPYVSRLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSDQLLNCWQIA	836
Db	771	AYVMAGVSPYVSRLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSDQLLNCWQIA	830
QY	837	KGMSYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDDIETEHADGGKVPWKMA	896
Db	831	KGMSYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDDIETEHADGGKVPWKMA	890
QY	897	LESILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID	956
Db	891	LESILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID	950
QY	957	VYIMVYKWMIDSECRPRELVSFSEFSEFSEFSEFSEFSEFSEFSEFSEFSEFSEFSEF	1016
Db	951	VYIMVYKWMIDSECRPRELVSFSEFSEFSEFSEFSEFSEFSEFSEFSEFSEFSEF	1010
QY	1017	DDMGDLVDAEYLVPOQGFCDPAPGAGGVMVHRRSSSTRSGGDLTLGLEPSEEA	1076
Db	1011	DDMGDLVDAEYLVPOQGFCDPAPGAGGVMVHRRSSSTRSGGDLTLGLEPSEEA	1070
QY	1077	PRSPAPSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP	1136
Db	1071	PRSPAPSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP	1130
QY	1137	LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERPCTLSPGKNGVYKDVAFGGA	1196
Db	1131	LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERPCTLSPGKNGVYKDVAFGGA	1190
QY	1197	VENPEYLTQGGAAPOHPHPPAFSPAFDNLVYWDQDPPBERGAPPSTFKGTPTAENPEYLG	1256
Db	1191	VENPEYLTQGGAAPOHPHPPAFSPAFDNLVYWDQDPPBERGAPPSTFKGTPTAENPEYLG	1250
QY	1257	LDVPV 1261	
Db	1251	LDVPV 1255	

RESULT 7
ID AAB21198
XX AAB21198 standard; protein; 1255 AA.
AC AAB21198;
DT 12-JAN-2001 (first entry)
XX Human HER-2/neu protein.
DE Human: HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.
XX Homo sapiens.
XX WO200044899-A1.
XX 03-AUG-2000.
PD 28-JAN-2000; 2000WO-US02164.
PF 29-JAN-1999; 99US-0117976.
XX

PA	(CORI-) CORIXA CORP.
PA	(SMIK) SMITHKLINE BEECHAM.
XX	Cheever MA, Gheysen D;
XX	WPI; 2000-505976/45.
DR	N-PSDB; AAA89736.
XX	HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT	useful for vaccinating against breast, ovarian, colon, lung and
PT	prostate cancers -
XX	Claim 52; Fig 7; 128pp; English.
PS	The present sequence is the human HER-2/neu protein. It is a member of
CC	the tyrosine kinase family of receptor-like glycoproteins and shows
CC	homology to the epidermal growth factor receptor (EGFR). It probably
CC	plays a part in cell growth and/or differentiation. The HER-2/neu
CC	gene is an oncogene. An HER-2/neu fusion protein comprising a
CC	HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC	domain may be used to treat or prevent cancer by eliciting or
CC	enhancing an immune response to the HER-2/neu protein. It may be used
CC	to treat malignancies such as breast, ovarian, colon, lung and
CC	prostate cancers, and may be used as an antigen to vaccinate against
CC	these neoplasias.
XX	Sequence . 1255 AA;
SQ	

Query Match	98.0%; Score 6708; DB 21; Length 1255;						
Best Local Similarity	98.1%; Pred. No. 0;						
Matches 1241; Conservative	1; Mismatches 9; Indels 14; Gaps 2;						
QY	1	MELAALCRWGLLALLPPGAAS	TGCTG	DMKLRLPAS	PTHLDMLRHLVQGCQVQGNL	60	
Db	1	MELAALCRWGLLALLPPGAAS	TGCTG	DMKLRLPAS	PTHLDMLRHLVQGCQVQGNL	60	
QY	61	ELTYLPTNASLSFLQDIQEVQGV	YLI	AHNQVQVPLQRLRIVRGTQLFEDNYALAVLDNG	120		
Db	61	ELTYLPTNASLSFLQDIQEVQGV	YLI	AHNQVQVPLQRLRIVRGTQLFEDNYALAVLDNG	120		
QY	121	DPLNNFNFTVFWLRVPKVSAS	----	HL	QLRSLTEILKGGVLIQRNPOLCYQDTILWK	176	
Db	121	DPLNN	-----	TPVT	GA	SPGGLRELQLRSLTEILKGGVLIQRNPOLCYQDTILWK	170
QY	177	DIFHKNNQALTLIDTNRSRACHPCSP	CKSGRCWSESSEDQSLTRTVCCAGCARCKGP	236			
Db	171	DIFHKNNQALTLIDTNRSRACHPCSP	CKSGRCWSESSEDQSLTRTVCCAGCARCKGP	230			
QY	237	LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTF	FSMPNPEGRYT	296			
Db	231	LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTF	FSMPNPEGRYT	290			
QY	297	FGASCVTACPNYVLTVDVGSCTLVCPHLNQEVTAEDGTORCEKSPCARVCYGLGMEHL	356				
Db	291	FGASCVTACPNYVLTVDVGSCTLVCPHLNQEVTAEDGTORCEKSPCARVCYGLGMEHL	350				
QY	357	REVRAVTSANIOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOEQLOVPETLEEITGYL	416				
Db	351	REVRAVTSANIOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOEQLOVPETLEEITGYL	410				
QY	417	YISAWPDSLPDLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGSGALAIHH	476				
Db	411	YISAWPDSLPDLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGSGALAIHH	470				
QY	477	NTHLCFVHTVPWDQLFRNPQHALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVN	536				
Db	471	NTHLCFVHTVPWDQLFRNPQHALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVN	530				
QY	537	CSQFLRGQECVEECRVLQGLPREYVNNARHCLPCHPECPQNGSVTCFGEADQCVACAHY	596				
Db	531	CSQFLRGQECVEECRVLQGLPREYVNNARHCLPCHPECPQNGSVTCFGEADQCVACAHY	590				
QY	597	KDPPFCVACPSGVKPDLSYMPYIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEORASP	656				

Db 591 KDPFFCVARCGVKPDLSTWPKFPDEEGACQPCINCTHSCVDLDDGCAEQRASP 650
Qy 657 LTSIVSAVGGILVWLVGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
Db 651 LTSIIISAVVGGILVWLVGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710
Qy 717 QMRILKETELURKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDE 776
Db 711 QMRILKETELURKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDE 770
Qy 777 AYVMAGVGSPPVSRLLGICLTSTVOLVTQMLPVCGLLDHVRNRRGLGSODLWNCQIA 836
Db 771 AYVMAGVGSPPVSRLLGICLTSTVOLVTQMLPVCGLLDHVRNRRGLGSODLWNCQIA 830
Qy 837 KGMYSILEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKWWA 896
Db 831 KGMYSILEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKWWA 890
Qy 897 LESILRRRFTHQSDVMSGYVTWELMTFGAKPYDGIIPAREIPDLEKGERLPPOPICTID 956
Db 891 LESILRRRFTHQSDVMSGYVTWELMTFGAKPYDGIIPAREIPDLEKGERLPPOPICTID 950
Qy 957 VYIMVWKWIMIDSECRPRFRELVSERFARMARDQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
Db 951 VYIMVWKWIMIDSECRPRFRELVSERFARMARDQRFVVIQNEDLGPASPLDSTFYRSLLE 1010
Qy 1017 DDDMGDLVDAEYLVFQQGFCDPAPAGGMVHRRSSSTRSGGDLTLGLEPSEEEA 1076
Db 1011 DDDMGDLVDAEYLVFQQGFCDPAPAGGMVHRRSSSTRSGGDLTLGLEPSEEEA 1070
Qy 1077 PRSPLAPSGAGSDVFDGLGMAAKGLOSPLTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSGAGSDVFDGLGMAAKGLOSPLTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130
Qy 1137 LTCSPQPEYVNOVDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGGA 1196
Db 1131 LTCSPQPEYVNOVDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGGA 1190
Qy 1197 VENPEYLTQGGAAPOHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTQGGAAPOHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 8

AA84780
ID AA84780 standard; Protein; 1255 AA.

XX AA84780;

AC AA84780;

DT 08-AUG-2000 (first entry)

DE Amino acid sequence of the SPLICE erbB-2 receptor protein.

XX SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.

OS Homo sapiens.

XX WO200020579-A1.

PN 13-APR-2000.

PD 01-OCT-1999; 99WO-CA00912.

XX 02-OCT-1998; 98US-0165192.

PR 02-OCT-1998; 98US-0165192.

XX 02-OCT-1998; 98US-0165192.

(UYMC-) UNIV MCMASTER.

XX Muller WJ, Siegel PM;

XX WPI: 2000-303768/26.

DR N-PSDB; AAA14812.

XX Nucleic acid encoding an erbB 2 receptor protein designated SPLICE

PT erbB-2, inhibitors of the protein are useful for treatment of cancer -

XX Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPLICE erbB-2 receptor protein. The

CC protein has an in-frame deletion of 16 amino acids, 2 of which are

CC conserved cysteine residues, compared to the unspliced protein. The

CC erbB-2 polynucleotide is used to construct probes for detecting

CC disorders of cell transformation such as cancer. Antibodies to the

CC protein may be used to detect SPLICE erbB-2 in a sample. Agents

CC (e.g. antisense oligonucleotides) which inhibit the expression of

CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and

CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful

CC for treating conditions involving damaged cells including conditions

CC in which degeneration of tissue occurs, such as arthropathy, bone

CC resorption, inflammatory diseases, degenerative disorders of the

CC central nervous system and wound healing.

XX SQ Sequence 1255 AA;

Query Match 98.0%; Score 6708; DB 21; Length 1255;

Best Local Similarity 98.1%; Pred. No. 0;

Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;

Qy 1 MELAALCRWGLLIALLPGAASQVCTGTDKMLRLPASPTHLDMLRHLVQGCVOVQGNL 60

Db 1 MELAALCRWGLLIALLPGAASQVCTGTDKMLRLPASPTHLDMLRHLVQGCVOVQGNL 60

Qy 61 ELTYLPTNASLSPLQIQEVQGVYLIQAHNQVROVPLQRLRIVRGTLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSPLQIQEVQGVYLIQAHNQVROVPLQRLRIVRGTLFEDNYALAVLDNG 120

Qy 121 DPLNNFNTVSPWLVKPVKVAS-----TTPVTGASPGGLRELRLSLTEILKGGVLIQRPOLCYQDTTLWK 176

Db 121 DPLNNFNTVSPWLVKPVKVAS-----TTPVTGASPGGLRELRLSLTEILKGGVLIQRPOLCYQDTTLWK 170

Qy 177 DIFHKNNQALTLIDTNRSRACHPCSPMCKSGCWSESSEDQSLTRTVCAGGCARCKGP 236

Db 171 DIFHKNNQALTLIDTNRSRACHPCSPMCKSGCWSESSEDQSLTRTVCAGGCARCKGP 230

Qy 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 296

Db 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290

Qy 297 FGASCVTACPNYVLTVDGSCITLVCPHNOEVTAEQTCRCEKSPCARVCYGLGMEHL 356

Db 291 FGASCVTACPNYVLTVDGSCITLVCPHNOEVTAEQTCRCEKSPCARVCYGLGMEHL 350

Qy 357 REVRAVTSANIOEFAGCKKIFGSLAPLPSFDGDPASNTAPLOEQLOVFTLEEITGYL 416

Db 351 REVRAVTSANIOEFAGCKKIFGSLAPLPSFDGDPASNTAPLOEQLOVFTLEEITGYL 410

Qy 417 YISAWPDSLPLDSVFNQVIRGRILHNGAYSILTLQGLGISWLGRLSRLGSLALIH 476

Db 411 YISAWPDSLPLDSVFNQVIRGRILHNGAYSILTLQGLGISWLGRLSRLGSLALIH 470

Qy 477 NTHLCFVHTVPWDOLFNRNPHQALLTANRPECEVCEGLACHQLCARGHCWGPGTQCVN 536

Db 471 NTHLCFVHTVPWDOLFNRNPHQALLTANRPECEVCEGLACHQLCARGHCWGPGTQCVN 530

Qy 537 CSQFLRGQECVEECRVLQGLPREYVNHARCLPCHPECPQNGSVTCFGEADOCVACAHY 596

Db 531 CSQFLRGQECVEECRVLQGLPREYVNHARCLPCHPECPQNGSVTCFGEADOCVACAHY 590

Qy 597 KDPFFCVARCGVKPDLSTWPKFPDEEGACQPCINCTHSCVDLDDGCAEQRASP 656

Db 591 KDPFFCVARCPGSKPDLSPYPIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEORASP 650
597 LTVSIVAVGILLVVLGVVFGILIKRQOKIRKTYMRRLLQETELVEPLTPSGAMPNQA 716
651 LTVSIVAVGILLVVLGVVFGILIKRQOKIRKTYMRRLLQETELVEPLTPSGAMPNQA 710
717 QMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKELDE 776
711 QMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKELDE 770
777 AYVMAGVSPYSRLLGICLTSTVQLVTPQMPYGCCLLDHVRNRRGLSGDOLLNWCQIA 836
771 AYVMAGVSPYSRLLGICLTSTVQLVTPQMPYGCCLLDHVRNRRGLSGDOLLNWCQIA 830
837 KMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKWA 896
831 KMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKWA 890
897 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTID 956
891 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTID 950
957 VYIMVWKCMIDSECRPRFRELVSFSRMDRQRFVIONEDLGPASPLDSTFYRSLLE 1016
951 VYIMVWKCMIDSECRPRFRELVSFSRMDRQRFVIONEDLGPASPLDSTFYRSLLE 1010
1017 DDDMGDLVDAEYLVPOQGFCDPAPGAGVMVHRRHSSTRSGGDLTLGLEPSEEEA 1076
1011 DDDMGDLVDAEYLVPOQGFCDPAPGAGVMVHRRHSSTRSGGDLTLGLEPSEEEA 1070
1077 PRSPLAPSEGAGSDVFDGLGMAAGLQSLTHDPSPLQYSEDPTVPLPSETDGYVAP 1136
1071 PRSPLAPSEGAGSDVFDGLGMAAGLQSLTHDPSPLQYSEDPTVPLPSETDGYVAP 1130
1137 LTCSPQPEYVNOQDVRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGA 1196
1131 LTCSPQPEYVNOQDVRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGA 1190
1197 VENPEYLTPOGGAAPQHPPPAFSPAFDNLYYDQDPPERGAPPSTFKGTPTAENPEYLG 1256
1191 VENPEYLTPOGGAAPQHPPPAFSPAFDNLYYDQDPPERGAPPSTFKGTPTAENPEYLG 1250
1257 LDVPV 1261
1251 LDVPV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX AAB85458;
XX DE
XX 25-SEP-2001 (first entry)
XX Human HER-2/neu protein.
XX Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
XX oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX Homo sapiens.
XX WO200153463-A2.
XX 26-JUL-2001.
XX 19-JAN-2001; 2001WO-US01850.
XX 21-JAN-2000; 2000US-0177545.
XX (CORI-) CORIXA CORP.
XX

PI Cheever MA, Hand-Zimmermann S;
XX WPI: 2001-476112/51.
DR N-PSDB; AAH23392.
XX
PT New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer -
XX
XX Claim 2; Page 41-46; 49pp; English.
XX
CC The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;

Query Match 98.0%; Score 6708; DB 22; Length 1255;
Best Local Similarity 98.1%; Pred. No. 0;
Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASTOVCTGTDMLRLPASPTHLDMLRHLVGGCVVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASTOVCTGTDMLRLPASPTHLDMLRHLVGGCVVQGNL 60

QY 61 ELTYLPTNASLSFLQIQEVQGVLIHQAQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQIQEVQGVLIHQAQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

QY 121 DPLNNFNTVSPWLRVVKVAS-----HLEQLRSLTEILKGGVLIQRNPOLCYQDTLLWK 176
DB 121 DPLNN-----TTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTLLWK 170

QY 177 DIFHKNNQALTLIDTNRSRACHPCSPCKSGRCWGESSEDCOSLRTVTCAGCARCKGP 236
DB 171 DIFHKNNQALTLIDTNRSRACHPCSPCKSGRCWGESSEDCOSLRTVTCAGCARCKGP 230

QY 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRT 296
DB 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRT 290

QY 297 FGASCVTACPNYLSLTDVGSCTLVCPHNOEVTAEQTCRCKSKPCARVCYGLGMEHL 356
DB 291 FGASCVTACPNYLSLTDVGSCTLVCPHNOEVTAEQTCRCKSKPCARVCYGLGMEHL 350

QY 357 REVRAVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVETLEEITGYL 416
DB 351 REVRAVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVETLEEITGYL 410

QY 417 YISAWPDSLPLDSVFONLQVIRGRILHNGAYSILTLQGLGISWLGSLRSLRGLSLALHH 476
DB 411 YISAWPDSLPLDSVFONLQVIRGRILHNGAYSILTLQGLGISWLGSLRSLRGLSLALHH 470

QY 477 NTHLCFVHTVPWDQLFRNPQALLHTANRPEBCEVGEGLACHOLCARGHCWGPGPTQCVN 536
DB 471 NTHLCFVHTVPWDQLFRNPQALLHTANRPEBCEVGEGLACHOLCARGHCWGPGPTQCVN 530

QY 537 CSQFLRGQECVBECEVLOGLPREYVNHARCLPCHPECOPQNGSVTCFGEADOCVACAHY 596
DB 531 CSQFLRGQECVBECEVLOGLPREYVNHARCLPCHPECOPQNGSVTCFGEADOCVACAHY 590

QY 597 KDPFFCVARCPGSKPDLSPYPIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEORASP 656
DB 591 KDPFFCVARCPGSKPDLSPYPIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEORASP 650

QY 657 LTSIVSIVAVGILLVVLGVVFGILIKRQOKIRKTYMRRLLQETELVEPLTPSGAMPNQA 716
DB 651 LTSIVSIVAVGILLVVLGVVFGILIKRQOKIRKTYMRRLLQETELVEPLTPSGAMPNQA 710

Db 471 NTHLCFVHTVPMDLFRNPHQALLHTANRDEBCEVGEGLACHOLCARGHCWGPGTQCVN 530
Qy 537 CSOFLRQECVEECRVLQGLPREYVVARHCLPCHPECPONGSVTCFGEADOCVACAHY 596
Db 531 CSOFLRQECVEECRVLQGLPREYVVARHCLPCHPECPONGSVTCFGEADOCVACAHY 590
Qy 597 KDPFFCVARCPGKVPDLSYMPYKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQRAS 656
Db 591 KDPFFCVARCPGKVPDLSYMPYKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQRAS 650
Qy 657 LTVSIVAVGILLVWLVGVVFGILIKRQOKIRKYMRRLLQETELVEPLTPSGAMPNOA 716
Db 651 LTVSIVAVGILLVWLVGVVFGILIKRQOKIRKYMRRLLQETELVEPLTPSGAMPNOA 710
Qy 717 QMRILKETELURKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKETLDE 776
Db 711 QMRILKETELURKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKETLDE 770
Qy 777 AYVMAGVGSYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSDLLNWCQIA 836
Db 771 AYVMAGVGSYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSDLLNWCQIA 830
Qy 837 KMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMA 896
Db 831 KMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMA 890
Qy 897 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 956
Db 891 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 950
Qy 957 VYIMVWKMMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1016
Db 951 VYIMVWKMMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1010
Qy 1017 DDMGDLVDAEELVLPQGGFCDDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEA 1076
Db 1011 DDMGDLVDAEELVLPQGGFCDDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEA 1070
Qy 1077 PRSPLAPSEGAGSDVFDGLGMAAGLQSLPLTHDPSPLQRYSEDDTVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGLGMAAGLQSLPLTHDPSPLQRYSEDDTVPLPSETDGYVAP 1130
Qy 1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVXKDVAFGGA 1196
Db 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVXKDVAFGGA 1190
Qy 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERCAPPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERCAPPSTFKGTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 11

AAE24067
ID AAE24067 standard; Protein; 1255 AA.

XX AC AAE24067;

XX DT 23-SEP-2002 (first entry)
XX DE Human Her-2 protein.

XX KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;

XX KW hyperproliferative disorder; prophylaxis; inflammation; anticise;

XX KW tumour; gene therapy; phosphorothioate backbone.

XX OS Homo sapiens.

XX PN W020022636-A1.

XX

PD 21-MAR-2002.
XX 12-SEP-2001; 2001WO-US28572.
XX 15-SEP-2000; 2000US-0663834.
XX (ISIS-) ISIS PHARM INC.
XX Bennett CF, Cowbert LM;
XX WPI; 2002-471192/50.
DR N-PSDB; AAD38904.
XX Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX Example 13; Page 95-107; 116pp; English.
XX The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX

SQ Sequence 1255 AA;

Query Match 98.0%; Score 6708; DB 23; Length 1255;
Best Local Similarity 98.1%; Pred No. 0;
Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;
Qy 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRPLASPETHLDMLRHLYGCGQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRPLASPETHLDMLRHLYGCGQVQGNL 60

Qy 61 ELTYLPNASLSFLDIOEQVGVYLAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPNASLSFLDIOEQVGVYLAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Qy 121 DPLNNFNFTVSWLRVPKVSAS- ---HLEQLASLTETILKGVLIQRNPOLCYODTILWK 176
Db 121 DPLNN- ---TTPVTGASPGGURELQRLSLETILKGVLIQRNPOLCYODTILWK 170
Qy 177 DIFHKNQALTLTIDNRSRACHPCSPMCKGSRGWGSESDCCSLTRTVCAAGGCARCKGP 236
Db 171 DIFHKNQALTLTIDNRSRACHPCSPMCKGSRGWGSESDCCSLTRTVCAAGGCARCKGP 230
Qy 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 296
Db 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 290
Qy 297 FGASCVTACPNYLTSTDVSGTLVCPLNHNSGICELHCPALVTYNTDTFESMPNPEGRYT 356
Db 291 FGASCVTACPNYLTSTDVSGTLVCPLNHNSGICELHCPALVTYNTDTFESMPNPEGRYT 350
Qy 357 REVRVTSANIQEFAGCKIFGSLAFLESFDGDPASNTAPLOPEQLQVFTLEETGYL 416
Db 351 REVRVTSANIQEFAGCKIFGSLAFLESFDGDPASNTAPLOPEQLQVFTLEETGYL 410
Qy 417 YISAWPDSLPLSVFQNLQVIRGRILHNGAYSILTQGLGSLWGLSLRSLRGLSLALHH 476
Db 411 YISAWPDSLPLSVFQNLQVIRGRILHNGAYSILTQGLGSLWGLSLRSLRGLSLALHH 470
Qy 477 NTHLCFVHTVPMDQPRNPHQALLHTANRDEBCEVGEGLACHOLCARGHCWGPGTQCVN 536
Db 471 NTHLCFVHTVPMDQPRNPHQALLHTANRDEBCEVGEGLACHOLCARGHCWGPGTQCVN 530
Qy 537 CSOFLRQECVEECRVLQGLPREYVVARHCLPCHPECPONGSVTCFGEADOCVACAHY 596

Db 351 REVRVTSANIQEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGYL 410
Qy 417 YISAWPDSLPLSVFQNLQVIRGRIILHNGAYSILTLOGIGISWILGRSLRSLGSLALIIH 476
Db 411 YISAWPDSLPLSVFQNLQVIRGRIILHNGAYSILTLOGIGISWILGRSLRSLGSLALIIH 470
Qy 477 NTHLCFVHTVPWDLFRNPHQALLHTANRDEDECVGEGGLACHQLCARGHCWGPOPTOCVN 536
Db 471 NTHLCFVHTVPWDLFRNPHQALLHTANRDEDECVGEGGLACHQLCARGHCWGPOPTOCVN 530
Qy 537 CSOFLRQECVEECRVLOGLPREYVYARHCLPCHPECPQNGSVTCFPGPADQVCAHY 596
Db 531 CSOFLRQECVEECRVLOGLPREYVYARHCLPCHPECPQNGSVTCFPGPADQVCAHY 590
Qy 597 KDPFPCVARGSPGVKPDLSYMPYKPFDEGACOPCPINCHSHCVOLDKGCRAEORASP 656
Db 591 KDPFPCVARGSPGVKPDLSYMPYKPFDEGACOPCPINCHSHCVOLDKGCRAEORASP 650
Qy 657 LTSIVSAVVGILLVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716
Db 651 LTSIVSAVVGILLVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710
Qy 717 QMRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDE 776
Db 711 QMRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDE 770
Qy 777 AYVMAGVSPVSRLLGILCTSTVOLVTOLMPYGCILLDHVRENHGRGLSDLLNWCQOIA 836
Db 771 AYVMAGVSPVSRLLGILCTSTVOLVTOLMPYGCILLDHVRENHGRGLSDLLNWCQOIA 830
Qy 837 KMSYLEDLVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKMA 896
Db 831 KMSYLEDLVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKMA 890
Qy 897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICTID 956
Db 891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICTID 950
Qy 957 YVIMVWKWIDSECRPRFRELSEFMRMDPQRFVYVIONEDLGASPLDSTFYRSILLE 1016
Db 951 YVIMVWKWIDSECRPRFRELSEFMRMDPQRFVYVIONEDLGASPLDSTFYRSILLE 1010
Qy 1017 DDDMGDLVDAEYLVLPQGGFCPPDPAPGAGMVHRRSSSTRSGGDLTLGLPSEEEA 1076
Db 1011 DDDMGDLVDAEYLVLPQGGFCPPDPAPGAGMVHRRSSSTRSGGDLTLGLPSEEEA 1070
Qy 1077 PRSLAPSEGAGSDVFDGDLGMGAAGKLSLPLTHDPSPLQRYSDPTVPLPSETDGYVAP 1136
Db 1071 PRSLAPSEGAGSDVFDGDLGMGAAGKLSLPLTHDPSPLQRYSDPTVPLPSETDGYVAP 1130
Qy 1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKDVFARFGA 1196
Db 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKDVFARFGA 1190
Qy 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAAPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAAPSTFKGTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 13

AAM51143

ID AAM51143 standard; Protein: 1255 AA.

XX

AC AAM51143;

XX

DT 17-JUN-2002 (first entry)

XX

DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX

KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KW tyrosine kinase; receptor; c-erbB2; gene therapy.
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT Domain 1..653
FT /note= "extracellular domain"
FT Domain 676..1255
FT /note= "intracellular domain"
FT Domain 990..1255
FT /note= "phosphorylation domain"
PN WO200212341-A2.
XX
XX
PD 14-FEB-2002.
XX
XX 03-AUG-2001; 2001WO-US24283.
XX
XX 03-AUG-2000; 2000US-0632507.
XX
XX (CORI-) CORIXA CORP.
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX Cheever MA, Gheysen D;
XX MPI; 2002-241743/29.
XX N-PSDB; ABA92250.
XX
XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
XX or enhancing an immune response to the protein, has Her-2/neu
XX extracellular domain fused to Her-2/neu intracellular or
XX phosphorylation domain -
XX
XX Claim 68; Fig 7; 141pp; English.
XX
XX The present sequence is that of human Her-2/neu (p185 glycoprotein
XX or c-erbB2), an oncogenic self-protein and target for anti-cancer
XX vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
XX in a variety of cancers, including breast, ovarian, colon, lung and
XX prostate cancer. Her-2/neu is a member of the tyrosine kinase
XX family of receptor-like glycoproteins. It comprises an extracellular
XX domain with homology to the epidermal growth factor receptor
XX (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
XX intracellular domain that also shows homology to EGFR. Its
XX overexpression correlates with a poor prognosis in breast and
XX ovarian cancers. The invention provides Her-2/neu fusion
XX proteins, nucleic acids encoding them, viral vectors, and vaccines
XX comprising the fusion proteins or nucleic acid molecules. In
XX preferred fusion proteins, the extracellular domain of a Her-2/neu
XX protein is fused to a Her-2/neu intracellular domain or
XX phosphorylation domain (or its DeltaPD fragment). An immune
XX response to Her-2/neu protein is elicited or enhanced by
XX administering the fusion protein in the form of a vaccine, or by
XX transfecting cells of an animal ex vivo with a nucleic acid
XX encoding the fusion protein, and delivering the transfected cells
XX to the animal. The fusion proteins, nucleic acids, and isolated
XX specific T-cells are useful for inhibiting the development of a
XX cancer, especially breast, ovarian, colon, lung or prostate cancer
XX in a patient. T cells that specifically react with a Her-2/neu
XX fusion protein can be used to remove tumour cells from a sample in
XX order to inhibit the development of cancer in a patient.
XX
XX Sequence 1255 AA;

Query Match

Best Local Similarity 98.0%; Score 6708; DB 23; Length 1255;

Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLRASPETHLDMLRHLHYGGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLRASPETHLDMLRHLHYGGCVVQGNL 60

QY 61 ELTYLPTNASLSFLQDIOEVQGVLIHAHNOVROVPLQRLRIVRGTOLFPEDNYALAVLDNG 120
DB |||||
QY 61 ELTYLPTNASLSFLQDIOEVQGVLIHAHNOVROVPLQRLRIVRGTOLFPEDNYALAVLDNG 120
DB |||||
QY 121 DFLNNFNFTVSWFLRPVKVSAS---HLEQKRSLTEILKGGVLIQRNPQLCYQDTILWK 176
DB |||||
QY 121 DFLNN-----TTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWK 170
DB |||||
QY 177 DIFHKNNQALTLIDTNRSRACHPCSPMCKGSRGWGESSEDCQSLTRTVCAAGCARCKGP 236
DB |||||
QY 171 DIFHKNNQALTLIDTNRSRACHPCSPMCKGSRGWGESSEDCQSLTRTVCAAGCARCKGP 230
DB |||||
QY 237 LPTDCHECOAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 296
DB |||||
QY 231 LPTDCHECOAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 290
DB |||||
QY 297 FGASCVTACPNYVSTLSDVGSCTLVCPPLHNQEVTAEDGTQRCBKCKPCARVCYGLGMEHL 356
DB |||||
QY 291 FGASCVTACPNYVSTLSDVGSCTLVCPPLHNQEVTAEDGTQRCBKCKPCARVCYGLGMEHL 350
DB |||||
QY 357 REVRAVTSANIOEFAGCKKIFGSLAPLPESFGDPPASNTAPLOPEQLQVFTELEITGYL 416
DB |||||
QY 351 REVRAVTSANIOEFAGCKKIFGSLAPLPESFGDPPASNTAPLOPEQLQVFTELEITGYL 410
DB |||||
QY 417 YISAMPDSLPDLSVFQNLQVIRGRILHNGAYSITLQGLGISLGLRLSRELGSGLAIHH 476
DB |||||
QY 411 YISAMPDSLPDLSVFQNLQVIRGRILHNGAYSITLQGLGISLGLRLSRELGSGLAIHH 470
DB |||||
QY 477 NTHLCFVHTVPWDQLFRNPQHALLHTANRPEBECVGEGLACHQLCARGCHWGPPTQCVN 536
DB |||||
QY 471 NTHLCFVHTVPWDQLFRNPQHALLHTANRPEBECVGEGLACHQLCARGCHWGPPTQCVN 530
DB |||||
QY 537 CSQFLRGQCEVBEICRVLQGLPREYVNAHCLPCHPECCQNGSVTCFGEPAOQCACAHY 596
DB |||||
QY 531 CSQFLRGQCEVBEICRVLQGLPREYVNAHCLPCHPECCQNGSVTCFGEPAOQCACAHY 590
DB |||||
QY 597 KDPFPCVACRPSGVKPDLSYMPIWKFPDDEGACQPCPINCTHSCVDLDDKGCAPORASP 656
DB |||||
QY 591 KDPFPCVACRPSGVKPDLSYMPIWKFPDDEGACQPCPINCTHSCVDLDDKGCAPORASP 650
DB |||||
QY 657 LTSIVSAVVGILLVVVLGVVFGILIKRRQOKIRKYMTRRLQETELVEPLTPSGAMPNOA 716
DB |||||
QY 651 LTSIVSAVVGILLVVVLGVVFGILIKRRQOKIRKYMTRRLQETELVEPLTPSGAMPNOA 710
DB |||||
QY 717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRNTPSKANKEILDE 776
DB |||||
QY 711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRNTPSKANKEILDE 770
DB |||||
QY 777 AYVMAGVSPYVSRLLGICLTSTVOLVTOLMPYGCLLDHVNRGRGLSGQDILNWCMTIA 836
DB |||||
QY 771 AYVMAGVSPYVSRLLGICLTSTVOLVTOLMPYGCLLDHVNRGRGLSGQDILNWCMTIA 830
DB |||||
QY 837 KGMSYLEDVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVPKMA 896
DB |||||
QY 831 KGMSYLEDVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVPKMA 890
DB |||||
QY 897 LESILRRRFTHSDVMSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 956
DB |||||
QY 891 LESILRRRFTHSDVMSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 950
DB |||||
QY 957 VYIMVVKCWMIDSECRPRFRELVSFESRWARDPQRVVIQNEIDLGPASPLDSTFYRSLLE 1016
DB |||||
QY 951 VYIMVVKCWMIDSECRPRFRELVSFESRWARDPQRVVIQNEIDLGPASPLDSTFYRSLLE 1010
DB |||||
QY 1017 DDDMGDLVDAEYLVPOQGFCCFPDPAAGAGGVHHRHRSSTRSGGDLTLGLEPSEEA 1076
DB |||||
QY 1011 DDDMGDLVDAEYLVPOQGFCCFPDPAAGAGGVHHRHRSSTRSGGDLTLGLEPSEEA 1070
DB |||||
QY 1077 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAP 1136
DB |||||
QY 1071 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAP 1130
DB |||||
QY 1137 LTCSPQPEYVNOQDVRPQPPSPREGPLPAARPAGATLERAKTSLSPGKNGVVKDVFAPGGA 1196

DB 1131 LTCSPQPEYVNOQDVRPQPPSPREGPLPAARPAGATLERAKTSLSPGKNGVVKDVFAPGGA 1190
QY 1197 VENPEYLTTPQGGAAPOPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
DB 1191 VENPEYLTTPQGGAAPOPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250
QY 1257 LDVPV 1261
DB 1251 LDVPV 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
XX
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX
XX WPI; 2002-280741/32.
DR
DR N-PSDB; ABK10730.
XX
XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide -
XX
PS Disclosure; Page 71-74; 74pp; English.
XX
CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 98.0%; Score 6708; DB 23; Length 1255;
Best Local Similarity 98.1%; Pred. No. 0;
Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;

QY 1 MELAAALCRWGLLALLPPCAASTQVCTGTDMLRLPASPETHLDMRLHLYQGCQVQGNL 60
DB |||||
QY 1 MELAAALCRWGLLALLPPCAASTQVCTGTDMLRLPASPETHLDMRLHLYQGCQVQGNL 60
DB |||||
QY 61 ELTYLPTNASLSFLQDIOEVQGVLIHAHNOVROVPLQRLRIVRGTOLFPEDNYALAVLDNG 120


```
Db 61 ELTYLPTNASLSFLQDIQEVGVVLIAHNOVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Qy 121 DPLANNFNFTVSWLRVPKVSAS-----HLEQLRSLTEILKGGVLIQORNPOLCYQDITLWK 176
Db 121 DPLNN-----TTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWK 170
Qy 177 DIFHKNNQALTLIDTNRSRACHPCSPMCKSGCWGESSEDCOSLTRTVCAGGCARCKGP 236
Db 171 DIFHKNNQALTLIDTNRSRACHPCSPMCKSGCWGESSEDCOSLTRTVCAGGCARCKGP 230
Qy 237 LPTDCHEQCAAGCTGPKHSDCIACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 296
Db 231 LPTDCHEQCAAGCTGPKHSDCIACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290
Qy 297 FGASCVTACPNYLSLTVDSGCTLVCPLNQOEVTAEQTCRCEKSPCARVCYGLGMEHL 356
Db 291 FGASCVTACPNYLSLTVDSGCTLVCPLNQOEVTAEQTCRCEKSPCARVCYGLGMEHL 350
Qy 357 REVRAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGYL 416
Db 351 REVRAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGYL 410
Qy 417 YISAWPDSLPDLVSFQNLQVIRGRIILHNGAYSITLQGLGISWGLRSLRELGSGLALIH 476
Db 411 YISAWPDSLPDLVSFQNLQVIRGRIILHNGAYSITLQGLGISWGLRSLRELGSGLALIH 470
Qy 477 NTHLCFVHTVPMDQLFRNPHQALLHTANRDEDCVGEGLACHOLCARGHCWGPQTQCVN 536
Db 471 NTHLCFVHTVPMDQLFRNPHQALLHTANRDEDCVGEGLACHOLCARGHCWGPQTQCVN 530
Qy 537 CSOFLRQECVEECRVLQGLPREYVNAHCLCPHCEQPONGSVTCFGEADOCVACAHY 596
Db 531 CSOFLRQECVEECRVLQGLPREYVNAHCLCPHCEQPONGSVTCFGEADOCVACAHY 590
Qy 597 KDPFFCVARCPGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEORASP 656
Db 591 KDPFFCVARCPGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEORASP 650
Qy 657 LTSISAVGILLVAVLVGVFGLIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQA 716
Db 651 LTSISAVGILLVAVLVGVFGLIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQA 710
Qy 717 QMRILKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPPKANKILDE 776
Db 711 QMRILKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPPKANKILDE 770
Qy 777 AYVMAGVGSPPYVSRLLIGICLTSTVQLVTQMPYGCILLDHVRENRLGSDLLNWCWQIA 836
Db 771 AYVMAGVGSPPYVSRLLIGICLTSTVQLVTQMPYGCILLDHVRENRLGSDLLNWCWQIA 830
Qy 837 KGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPWKMA 896
Db 831 KGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPWKMA 890
Qy 897 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTID 956
Db 891 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTID 950
Qy 957 VYMIWKWMIIDSECPRRFRELVSFPRMARDQRFVVIQNEIDLGPASPLDSTFYRSLLE 1016
Db 951 VYMIWKWMIIDSECPRRFRELVSFPRMARDQRFVVIQNEIDLGPASPLDSTFYRSLLE 1010
Qy 1017 DDDMGDLVDAEYLVPOQGFCCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEA 1076
Db 1011 DDDMGDLVDAEYLVPOQGFCCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEA 1070
Qy 1077 PRSPLAPSEGAGSDVFDGLGMGAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGLGMGAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1130
Qy 1137 LTCSPQPEYVNOVDVRPQPPSPREGPLPAARPAAGATLERKTLSPGKNGVVKDVFAGGA 1196
```

```
Db 1131 LTCSPQPEYVNOVDVRPQPPSPREGPLPAARPAAGATLERKTLSPGKNGVVKDVFAGGA 1190
Qy 1197 VENPEYLTPOGGNAOPHPHPAPFSPAFDNLTYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGNAOPHPHPAPFSPAFDNLTYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN WO9316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU ) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Houston LL, Huston JS, Oppermann H, Ring DB;
XX
WPI: 1993-272889/34.
DR N-PSDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

Query Match 97.3%; Score 6665; DB 14; Length 1433;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1234; Conservative 3; Mismatches 14; Indels 14; Gaps 2;

Qy 1 MELAALCRNGLLALLPFGAASQVCTGDMKRLPASPTHLDMLRHLYQGVQVQGNL 60
Db 1 MELAALCRNGLLALLPFGAASQVCTGDMKRLPASPTHLDMLRHLYQGVQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVGVVLIAHNOVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVGVVLIAHNOVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Qy 121 DPLANNFNFTVSWLRVPKVSAS-----HLEQLRSLTEILKGGVLIQORNPOLCYQDITLWK 176
Db 121 DPLNN-----TTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWK 170
Qy 177 DIFHKNNQALTLIDTNRSRACHPCSPMCKSGCWGESSEDCOSLTRTVCAGGCARCKGP 236
Db 171 DIFHKNNQALTLIDTNRSRACHPCSPMCKSGCWGESSEDCOSLTRTVCAGGCARCKGP 230
```


A:Cross-references: GB:M16792; NID:q183983; PIDN:AAA58637.1; PID:q553332
C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C:Genetics:
A:Gene: GDB:ERBB2; NGL; NEU; HER-2
A:Cross-references: GDB:120613; OMIM:164870
A:Map position: 17q21.1-17q21.1
A:Introns: 25/1; 75/3; 147/1; 883/3
A:Note: The list of introns is incomplete
C:Function:
A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F:22-653/Domain: extracellular #status predicted <EXT>
F:70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F:654-675/Domain: transmembrane #status predicted <TM>
F:676-1255/Domain: intracellular #status predicted <INT>
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif
F:68,124,187,259,530,571,629/Binding site: carboxydrate (Asn) (covalent) #status predicted
F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:753/Active site: Lys #status predicted
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.0%; Score 6709; DB 1; Length 1255;
Best Local Similarity 98.0%; Pred. No. 2.9e-266;
Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;

QY 1 MELAALCRWGLLLALLPGCAASTOVCTGDMKRLPASPTHLDMLRHLYQCGVQGNL 60
DB 1 MELAALCRWGLLLALLPGCAASTOVCTGDMKRLPASPTHLDMLRHLYQCGVQGNL 60

QY 61 EUTYLPNTASLSFLQDIOEQVGVLIHNRQVPLQRLRIRVGTQLFEDNVALAVLDNG 120
DB 61 EUTYLPNTASLSFLQDIOEQVGVLIHNRQVPLQRLRIRVGTQLFEDNVALAVLDNG 120

QY 121 DPLNNTPTVTGASPGSLRQLRLSLEILKGGVLIQNPOLCYQDTILMKDF---NNFTV 177
DB 121 DPLNNTPTVTGASPGSLRQLRLSLEILKGGVLIQNPOLCYQDTILMKDF---NNFTV 177

QY 178 SEWLRVPKVSASHLE-NRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKGP 236
DB 178 -----QALTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKGP 230

QY 237 LPTDCHEQCAAGCTGPKHSDCLACLFHNSHIGICELHCPALVTYNTDTFESMNPGRYT 296
DB 231 LPTDCHEQCAAGCTGPKHSDCLACLFHNSHIGICELHCPALVTYNTDTFESMNPGRYT 290

QY 297 FGASCVTACPNYLSLTDVGSCTLVCPLNHOEVTAEDGTORCKSKPCARVCYGLGMEHL 356
DB 291 FGASCVTACPNYLSLTDVGSCTLVCPLNHOEVTAEDGTORCKSKPCARVCYGLGMEHL 350

QY 357 REVRAVTSANIOEFACKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFLEETITGYL 416
DB 351 REVRAVTSANIOEFACKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFLEETITGYL 410

QY 417 YISAMPDLSLPSVFNQLQVIRGRILHNGAYSLSLTQGLGISWGLRSLRELGSGLAIHH 476
DB 411 YISAMPDLSLPSVFNQLQVIRGRILHNGAYSLSLTQGLGISWGLRSLRELGSGLAIHH 470

QY 477 NTHLCFVHTVPDOLFRNPHQALLHTANRPEDECVGEGLAGHOLCARGHCWGPGPTQCVN 536
DB 471 NTHLCFVHTVPDOLFRNPHQALLHTANRPEDECVGEGLAGHOLCARGHCWGPGPTQCVN 530

QY 537 CSQFLRGQCEBECRVLQGLPREYNARHCLPCHPCQPNQSVTCFGEADQCVACAHY 596
DB 531 CSQFLRGQCEBECRVLQGLPREYNARHCLPCHPCQPNQSVTCFGEADQCVACAHY 590

QY 597 KDPPFCVAPCSGVKPDLSYMTWPKFDPDEGACQPCPINCTHSCVDLDDKGCPAEQRAS 656

Db 591 KDPPFCVAPCSGVKPDLSYMTWPKFDPDEGACQPCPINCTHSCVDLDDKGCPAEQRAS 650
QY 657 LTSIVSAVVGILLVVLGVVFGILLKRRQOKIRKYMRLLOETELVELPTTSGAMPNOA 716
Db 651 LTSIIISAVVGILLVVLGVVFGILLKRRQOKIRKYMRLLOETELVELPTTSGAMPNOA 710
QY 717 QMRILKETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDE 776
Db 711 QMRILKETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDE 770
QY 777 AYVMAVGSPPVSRLLGICLTSTVOLVTQLMYPGCLLDHVRNRRGLSGQDLLNMCQIA 836
Db 771 AYVMAVGSPPVSRLLGICLTSTVOLVTQLMYPGCLLDHVRNRRGLSGQDLLNMCQIA 830
QY 837 KGMSTYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKMA 896
Db 831 KGMSTYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKMA 890
QY 897 LESILRRRFTHOSDVMSYGVTVWELMTFCAGKPYDGIIPAREIPDLLEKGERLPQPCTID 956
Db 891 LESILRRRFTHOSDVMSYGVTVWELMTFCAGKPYDGIIPAREIPDLLEKGERLPQPCTID 950
QY 957 VYIMVVKCWMIDSECRPRFRELVSFMRARDPQRFVWIONEDLGPASPLDSTFYRSLLE 1016
Db 951 VYIMVVKCWMIDSECRPRFRELVSFMRARDPQRFVWIONEDLGPASPLDSTFYRSLLE 1010
QY 1017 DDMGDLVDAEYLYVPOQGFCCPDAPGAGGVMVHRHSSSTRSGGDLTLGLEPSEEA 1076
Db 1011 DDMGDLVDAEYLYVPOQGFCCPDAPGAGGVMVHRHSSSTRSGGDLTLGLEPSEEA 1070
QY 1077 PRSPLAPSEGAGSDVFDGDLGMAAGLQSLTPHPSPLQRYSEDPTVLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGDLGMAAGLQSLTPHPSPLQRYSEDPTVLPSETDGYVAP 1130
QY 1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVKDVFAGGA 1196
Db 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERPKTSLSPGKNGVKDVFAGGA 1190
QY 1197 VENPEYLYPQGGAAQPPHPPAFSAFONLYYWDQDPPRGAPPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLYPQGGAAQPPHPPAFSAFONLYYWDQDPPRGAPPSTFKGTPTAENPEYLG 1250
QY 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 2
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) new precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C:Accession: A24562; A61204
R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A:Title: The new oncogene encodes an epidermal growth factor receptor-related protein
A:Reference number: A24562; MUID:86118662; PMID:3945311
A:Accession: A24562
A:Molecule type: mRNA
A:Residues: 1-1260 <BAR>
A:Cross-references: EMBL:X03362; NID:q56745; PIDN:CAA27059.1; PID:q56746
R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohe
Carcinogenesis 12, 1975-1978, 1991
A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals n
2-thiazolylformamide or N-methyl-N-nitrosourea.
A:Reference number: A61204; MUID:92035293; PMID:1682063
A:Accession: A61204
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 637-663, 'V', 665-702 <MAS>
A:Note: authors translated the codon GCA for residue 25 as Val
C:Genetics:
A:Gene: neu